59

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/63, 15/90, 15/82

A2

(11) International Publication Number: WO 95/15389

(43) International Publication Date: 8 June 1995 (08.06.95)

(21) International Application Number: PCT/NO94/00190

(22) International Filing Date: 24 November 1994 (24.11.94)

(30) Priority Data:
9324707.0 2 December 1993 (02.12.93) GB

(71)(72) Applicants and Inventors: OLSEN, Odd-Arne [NO/NO]; Tårnveien 16, N-1430 Ås (NO). KALLA, Roger [AU/AU]; Mowle Place, Weetangera, ACT 2614 (AU).

(74) Agent: A/S BERGEN PATENTKONTOR; Strandgaten 191, N-5000 Bergen (NO).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LT, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: PROMOTER

(57) Abstract

The *in vivo* expression in the aleurone cells of a cereal of a conjugate is described. The conjugate comprises a GOI (gene of interest) and a particular Ltp (lipid transfer protein) promoter - namely the Ltp2 gene promoter. The conjugate is stably integrated within the cereal's genomic DNA.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

PROMOTER

The present invention relates to a promoter and to a conjugate comprising the same. The present invention also relates to the use of the promoter for stage- and tissue- specific expression of a gene of interest (GOI). The present invention also relates to the genomic nucleotide sequence of, and isolation of, the promoter.

In particular the present invention relates to a promoter for a lipid transfer protein (Ltp) gene known as the Ltp2 gene. The present invention also relates to the application of this Ltp2 gene promoter to express a GOI specifically in the aleurone layer of a monocotyledon - especially a transgenic cereal seed - more especially a developing transgenic cereal seed.

A mature cereal seed contains two distinct organs: the embryo - which gives rise to the vegetative plant - and the endosperm - which supports the growth of the emerging seedling during a short period of time after germination. The endosperm, which is the site of deposition of different storage products such as starch and proteins, is further sub-divisible into a peripheral layer of living aleurone cells surrounding a central mass of non-living starchy endosperm cells.

The aleurone cells differentiate from primary endosperm cells early during seed development or between 10 to 21 days after fertilization. The aleurone layer and embryo share many similarities in their gene expression programmes. They are the only cereal seed tissues that survive the desiccation process during seed maturation and they both have active gene transcription during seed germination.

The aleurone layer of cereal seeds comprises specialized cells that surround the central starchy endosperm, i.e. the site for starch and protein accumulation in the developing seed (Bosnes *et al.*, 1992, Olsen *et al.*, 1992). During seed germination, the cells of the aleurone layer produce amylolytic and proteolytic enzymes that degrade the storage compounds into metabolites that are taken up and are used by the growing embryo. Two aspects of aleurone cell biology that have been intensively studied are the genetics of anthocyanin pigmentation of aleurone cells in maize (McClintock, 1987) and the hormonal regulation of gene transcription in the aleurone layer of germinating barley seeds (Fincher, 1989).

Using transposon tagging, several structural and regulatory genes in the anthocyanin synthesis pathway have been isolated and characterized (Paz-Ares *et al.*, 1987; Dellaporta *et al.*, 1988). In barley, alpha-amylase and beta-glucanase genes that are expressed both in the aleurone layer and embryos of mature germinating seeds have been identified (Karrer *et al.*, 1991; Slakeski and Fincher, 1992). In addition, two other cDNAs representing transcripts that are differentially expressed in the aleurone layers of developing barley grains have been isolated. These are CHI26 (Lea *et al.*,1991) and pZE40 (Smith *et al.*, 1992). For none of these gene products has it been shown in transgenic cereal plants that the promoter directs expression in just the aleurone layer of developing grains.

Non-specific lipid transfer proteins (nsLtp's) have the ability to mediate *in vitro* transfer of radiolabelled phospholipids from liposomal donor membranes to mitochondrial acceptor membranes (Kader *et al.*, 1984; Watanabe and Yamada, 1986). Although their *in vivo* function remains unclear, nsLTPs from plants have recently received much attention due to their recurrent isolation as cDNA clones representing developmentally regulated transcripts expressed in several different tissues. A common feature is that, at some point in development, they are highly expressed in tissues producing an extracellular layer rich in lipids. Thus, transcripts corresponding to cDNAs encoding 10 kDa nsLTPs have been characterized in the tapetum cells of anthers as well as the epidermal layers of leafs and shoots in tobacco (Koltunow *et al.*, 1990; Fleming *et al.*, 1992), and barley aleurone layers (Mundy and Rogers, 1986; Jakobsen *et al.*, 1989).

In addition, a 10 kDa nsLTP was discovered to be one of the proteins secreted from auxin-treated somatic carrot embryos into the tissue culture medium (Sterk *et al.*, 1991). Based on *in situ* data demonstrating that the Ltp transcripts are localized in the protoderm cells of the somatic and zygotic carrot embryo and in the epithelial layer of the maize embryonic scutellum, it was suggested that *in vivo* nsLTPs are involved in either cutin biosynthesis or in the biogenesis and degradation of storage lipids (Sossountzov *et al.*, 1991; Sterk *et al.*, 1991).

A nsLTP in *Arabidopsis* has been localized to the cell walls lending further support to an extracellular function if this class of proteins (Thoma *et al.*, 1993).

PCT WO 90/01551 mentions the use of the aleurone cells of mature, germinating seeds to produce proteins from GOIs under the control of an alpha-amylase promoter. This promoter is active only in germinating seeds.

Recently, using a standard *in vitro* Ltp assay, two 10 kDa and one member of a novel class of 7 kDa nsLtp's were isolated from wheat seeds (Monnet, 1990; Dieryck *et al.*, 1992). The sequence of this 7 kDa wheat nsLtp protein shows a high degree of similarity with the predicted protein from the open reading frame (ORF) of the Bz11E cDNA, which had been isolated in a differential screening for barley aleurone specific transcripts (Jakobsen *et al.*, 1989). However, the amino acid sequence of this polypeptide showed only limited sequence identities with the previously sequenced 10 kDa proteins. In sub-cellular localisation studies using gold labelled antibodies one 10 kDa protein from Arabidopsis was localised to the cell wall of epidermal leaf cells. The presence of a signal peptide domain in the N-terminus of the open reading frames of all characterised plant ns-LTP cDNAs, also suggests that these are proteins destined for the secretory pathway with a possible extracellular function.

Olsen *et al.* in a paper titled "Molecular Strategies For Improving Pre-Harvest Sprouting Resistance In Cereals" published in 1990 in the published extracts from the Fifth International Symposium On Pre-Harvest Sprouting In Cereals (Westview Press Inc.) describe three different strategies for expressing different "effector" genes in the aleurone layer and the scutellum in developing grains of transgenic plants. This document mentions 4 promoter systems - including a system called B11E (which is now recognised as being the same as the Ltp2 gene promoter). There is no sequence listing for B11E given in this document.

Kalla *et al.* (1993) in a paper titled "Characterisation of Promoter Elements Of Aleurone Specific Genes From Barley" describe the possibility of the expression of anti-sense genes by the use of promoters of the aleurone genes B22E, B23D, B14D, and B11E (which is now shown to be the same as Ltp2).

The Kalla *et al.* (1993) paper gives a very general map of the Ltp2 gene promoter. The transient expression results showed very low levels of expression of the reporter gene.

A sequence listing of the Ltp2 gene was available as of 23 December 1992 on the EMBL database.

One of the major limitations to the molecular breeding of new varieties of crop plants with aleurone cells expressing GOIs is the lack of a suitable aleurone specific promoter.

At present, the available promoters - such as the CaMV 35S, rice actin and maize alcohol dehydrogenase - all are constitutive. In this regard, they are non-specific in target site or stage development as they drive expression in most cell types in the plants.

Another problem is how to achieve expression of a product coded for by a GOI in the aleurone layer of the endosperm that gives minimal interference with the developing embryo and seedling.

It is therefore desirable to provide aleurone specific expression of GOIs in cereal such as rice, maize, wheat, barley and other transgenic cereal plants.

Moreover it is desirable to provide aleurone specific expression that does not lead to the detriment of the developing embryo and seedling.

According to a first aspect of the present invention there is provided a Ltp2 gene promoter comprising:

the sequence shown as SEQ. I.D. 1, or

a sequence that has substantial homology with that of SEQ. I.D. 1, or

a variant thereof.

According to a second aspect of the present invention there is provided a conjugate comprising a GOI and a Ltp2 gene promoter as just defined.

According to a third aspect of the present invention there is provided an in vivo expression system comprising a conjugate comprising a GOI and a Ltp2 gene promoter as just defined wherein the conjugate is integrated, preferably stably integrated, within a monocotyledon's (preferably a cereal's) genomic DNA.

According to a fourth aspect of the present invention there is provided a transgenic cereal comprising a conjugate comprising a GOI and a Ltp2 gene promoter as just defined wherein the conjugate is integrated, preferably stably integrated, within a cereal's genomic DNA.

According to a fifth aspect of the present invention there is provided the in vivo expression in the aleurone cells of a monocotyledon (preferably a cereal) of a conjugate comprising a GOI and a Ltp2 gene promoter as just defined; wherein the conjugate is integrated, preferably stably integrated, within the monocotyledon's genomic DNA.

According to a sixth aspect of the present invention there is provided a method of enhancing the in vivo expression of a GOI in just the aleurone cells of a monocotyledon (preferably a cereal) which comprises stably inserting into the genome of those cells a DNA conjugate comprising a Ltp2 gene promoter as just defined and a GOI; wherein in the formation of the conjugate the Ltp2 gene promoter is ligated to the GOI in such a manner that each of the myb site and the myc site in the Ltp2 gene promoter is maintained substantially intact.

According to a seventh aspect of the present invention there is provided the use of a myb site and a myc site in an Ltp2 gene promoter to enhance in vivo expression of a GOI in just in the aleurone cells of a monocotyledon (preferably a cereal) wherein the Ltp2 gene promoter and the GOI are integrated into the genome of the monocotyledon.

According to an eighth aspect of the present invention there is provided a method of enhancing the in vivo expression of a GOI in just the aleurone cells of a monocotyledon (preferably a cereal) which comprises stably inserting into the genome of those cells a DNA conjugate comprising a Ltp2 gene promoter as just defined and a GOI; wherein in the formation of th conjugate the Ltp2 gene promoter is ligated to the GOI in such a manner that any ne of the Sph1 site, the AL site or the DS site in the Ltp2 gene promoter is (are)

maintained substantially intact. The Sph1 site, the AL site and the DS site are defined later.

Preferably the promoter is a barley aleurone specific promoter.

Preferably the promoter is for a 7 kDa lipid transfer protein.

Preferably the promoter is used for expression of a GOI in a cereal seed.

Preferably the promoter is used for expression of a GOI in a monocotyledonous species, including a grass - preferably a transgenic cereal seed.

Preferably the cereal seed is anyone of a rice, maize, wheat, or barley seed.

Preferably the promoter is the promoter for Ltp2 of *Hordeum vulgare*.

Preferably at least one additional sequence is attached to the promoter gene or is present in the conjugate to increase expression of a GOI or the GOI.

The additional sequence may be one or more repeats (e.g. tandem repeats) of the promoter upstream box(es) which are responsible for the aleurone specific pattern of expression of Ltp2. The additional sequence may even be a *Sh1*-intron.

The term "GOI" with reference to the present invention means any gene of interest - but not the remainder of the natural Ltp2 gene for the cereal in question. A GOI can be any gene that is either foreign or natural to the cereal in question.

Typical examples of a GOI include genes encoding for proteins giving for example added nutritional value to the seed as a food or crop or for example increasing pathogen resistance. The GOI may even be an antisense construct for modifiying the expression of natural transcripts present in the relevant tissues.

Preferably the GOI is a gene encoding for any one of a protein having a high nutritional value, a *Bacillus thuringensis* insect toxin, or an alpha- or beta- amylase or germination induced protease antisense transcript.

The term "a variant thereof" with reference to the present invention means any substitution of, variation of, modification of, replacement of, deletion of or the addition of one or more nucleic acid(s) from or to the listed promoter sequence providing the resultant sequence exhibits aleurone specific expression.

The term "substantial homology" covers homology with respect to at least the essential nucleic acids of the listed promoter sequence providing the homologous sequence exhibits aleurone specific expression. Preferably there is at least 80% homology, more preferably at least 90% homology, and even more preferably there is at least 95% homology with the listed promoter sequence.

The term "maintained substantially intact" means that at least the essential components of each of the myb site and the myc site remain in the conjugate to ensure aleurone specific expression of a GOI. Preferably at least about 75%, more preferably at least about 90%, of the myb or myc site is left intact.

The term "conjugate", which is synonymous with the terms "construct" and "hybrid", covers a GOI directly or indirectly attached to the promoter gene to from a Ltp2-GOI cassette. An example of an indirect attachment is the provision of a suitable spacer group such as an intron sequence, such as the *Sh1*-intron, intermediate the promoter and the GOI.

The present invention therefore provides the novel and inventive use of an aleurone specific promoter - namely the use of the Ltp2 gene promoter, preferably the Ltp2 gene promoter from barley.

The main advantage of the present invention is that the use of the Ltp2 gene promoter results in specific aleurone expression of a GOI in the aleurone layer(s) of cereals such as rice, maize, wheat, barley and other transgenic cereal seeds, preferably maize seed.

It is particularly advantageous that the expression is both stage- and tissue- specific.

A further advantage is that the expression of the product coded for by a GOI in the aleurone layer of the endosperm gives minimal interference with the developing embryo and seedling. This is in direct contrast to constitutive promoters which give high levels of expression in the developing seedling and mature plant tissues which severely affect normal plant development.

The present invention is particularly useful for expressing GOI in the aleurone layer of developing grains - such as cereal seeds.

With regard to the present invention it is to be noted the EMBL database sequence listing *(ibid)* does not suggest that the Ltp2 gene promoter could be used to express a GOI in a stage- and tissue- specific manner. Also the database extract does not mention the importance of the myb gene segment or the myc gene segment.

It is also to be noted the paper titled "Molecular Strategies For Improving Pre-Harvest Sprouting Resistance In Cereals" *(ibid)* does not give any specific sequence listing information for the Ltp2 gene promoter. Also there is no explicit mention in this paper of using just the Ltp2 gene promoter to induce expression in just aleurone cells. Moreover, there is no mention in this paper of an Ltp2 - GOI conjugate being formed. Also there is no mention in this paper of the importance of the myb site or the myc site.

It is also to be noted that in the paper titled "Characterisation of Promoter Elements Of Aleurone Specific Genes From Barley" *(ibid)* there is no mention of an Ltp2 - GOI conjuagte stably integrated into genomic DNA of a cereal. Also there is no explicit disclosure of an <u>in vivo</u> expression system. Moreover, there is no full sequence listing in this paper for the Ltp2 gene promoter. Also there is no explicit mention in this paper of the importance of the myb site or the myc site of Ltp2 gene promoter for <u>in vivo</u> GOI expression.

In contrast to the work disclosed in PCT WO 90/01551, the Ltp2 gene promoter (which is not disclosed in PCT WO 90/01551) the Ltp2 gene promoter results in aleurone specific expression in developing grains.

In general, therefore, the present invention relates to a promoter for a Ltp2 gene encoding a 7 kDa nsLTP. *In situ* hybridization analysis demonstrates that the Ltp2 transcript is expressed exclusively in aleurone cells from the beginning of the differentiation stage and half-way into the maturation stage. Further commentary on the maturation stages is provided by Bosnes *et al.*, 1992.

The Ltp2 gene promoter may be inserted into a plasmid. For example, the Ltp2 BglII 0.84 kb fragment can be inserted into the *BamHI* site of Bluescript. A GOI, such as *GUS*, can then be inserted into this conjugate (construct). Furthermore, a *Sh*1 intron can then be inserted into the *Sma*I site of this conjugate.

Stable integration may be achieved by using the method of Shimamoto (1989). Another way is by bombardment of an embryonic suspension of cells (e.g. maize cells). Another way is by bombardment of immature embryos (e.g. barley embryos).

With the present invention, it can be shown by using particle bombardments that the -807 bp Ltp2 gene promoter fused to a beta-glucuronidase (*GUS*) reporter gene (which serves as a GOI) is active in the aleurone layer of developing barley seeds, giving 5% of the activity of the strong constitutive actin-promoter from rice. Also, in transgenic rice plants, the barley Ltp2-promoter directs strong expression of the *GUS*-reporter gene exclusively in the aleurone layer of developing seeds, suggesting the presence of conserved mechanisms for aleurone cell gene expression in the cereals.

In a preferred embodiment, the Ltp2 gene encodes a 7kDa barley seed nsLTP and has about 80% identity to the wheat 7kDa protein.

The transcript of the Ltp2 gene is detectable in the earliest morphologically distinguishable aleurone cells and accumulates during the differentiation stage to decline finally during seed maturation. It can also serve as a molecular marker for the differentiating aleurone cells as shown *in situ* hybridisation experiments where the spatial distribution of the transcript is to be examined.

In the present invention, a genomic clone was isolated using the cDNA insert of previously isolated cDNA clone pBz11E and characterised by DNA sequencing.

The sequence of the cDNA and isolated genomic clone was found to be identical in the overlapping region. It was found the Ltp2 gene does not contain any intron.

To prove that this is an active gene, the 5' region carried on a 845 bp DNA fragment delineated by two *Bgl* II restriction sites was fused to the *GUS* gene (following Jefferson 1987) and the construct was introduced into barley aleurone layers using micro projectile bombardment. Aleurone cells expressing *GUS* activity were detected proving that the gene promoter was indeed capable of driving the expression of the GOI in the relevant tissue.

By comparing the DNA sequence of this active promoter sequences several putative *cis*-acting elements with the potential of binding known transcriptional factors present in cereal aleurone layers were detected. They include the binding sites for transcriptional factors of the myb and myc class, namely TAACTG and CANNTG respectively. Our experiments showed that the myb and myc sites were important for good levels of expression.

Gel retardation experiments showed that the Ltp2 gene promoter has a myb site that is recognised by a MYB protein (e.g. from chicken).

In the present invention, mature fertile rice plants were regenerated from transformed cultured rice protoplasts. The developing seeds of these primary transformants were analysed for the expression of *GUS*. It was found that the barley seed Ltp2 gene promoter confers aleurone specific expression in transgenic rice plants. This is the first example of an aleurone specific promoter in developing seeds of a transgenic cereal.

The following were deposited in accordance with the Budapest Treaty at the recognised depositary The National Collections of Industrial and Marine Bacteria Limited (NCIMB) at 23 St Machar Drive, Aberdeen, Scotland, UK, AB2 1RY, on 22 November 1993:

(i) An *E. Coli* K12 bacterial stock containing the plasmid pLtp2pr - i.e. Bluescript containing the Ltp2 gene promoter (Deposit Number NCIMB 40598).

[To form pLtp2pr, the Ltp2 promoter of Figure 2b (see later) contained on a BglII fragment was inserted in the Bluescript KS vector into the BamHI site.]

(ii) An *E. Coli* K12 bacterial stock containing the plasmid pLtp2/GN - i.e. Bluescript containing a Ltp2 gene promoter - *GUS* conjugate (Deposit Number NCIMB 40599).

[To form pLtp2/GN, the *GUS*-reporter gene cassette (GN) contained on the SmaI-EcoRI fragment of the commercially available vector pBI101 (Clontech Inc.) was cloned directionally into the SmaI and EcoRI sites of pLtp2pr.]

(iii) An *E. Coli* K12 bacterial stock containing the plasmid pLtp2ΔBCIGN - i.e. Bluescript containing an Ltp2 gene promoter with a deletion spanning the myb and myc sites - *GUS* conjugate (Deposit Number NCIMB 40601).

[To form pLtp2ΔBCIGN, the Ltp2 promoter and the GN gene was inserted as described for pLtp2pr and pLtp2/GN except for the use of Bluescript SK and that the Ltp2 promoter was deleted in the myb-myc region (using a PCR strategy) as explained in the legend of Figure 7 (see later).]

(iv) An *E. Coli* K12 bacterial stock containing the plasmid pLtp2*Sh*1/GN - i.e. Bluescript containing an Ltp2 gene promoter-Sh1 intron-*GUS* conjugate (Deposit No. NCIMB 40600).

[To form pLtp2*sh*1/GN, the Ltp2 promoter and the GN gene was inserted as described for pLtp2pr and pLtp2/GN except for the use of Bluescript SK. The *Sh*1 intron from maize contained on a *Hinc*II restriction fragment was inserted into the SmaI site of this construct.]

Other embodiments and aspects of the present invention include: A transformed host having the capability of expressing a GOI in just the aleurone layer; A vector incorporating a conjugate as hereinbefore described or any part thereof; A plasmid comprising a conjugate as hereinbefore described or any part thereof; A cellular organism or cell line transformed with such a vector; A monocotylenedonous plant comprising any one of the same; A developping seed comprising any of the same; and A method of expressing any one of the same.

The present invention will now be described only by way of examples in which reference shall be made to the accompanying Figures in which:

Figure 1 is a nucleotide sequence of the Ltp2 gene;

Figure 2a is a nucleotide sequence of the Ltp2 gene promoter;
Figure 2b is a nucleotide sequence of the Ltp2 gene promoter with an additional 39 nucleotides for fusion to a *GUS* gene for transgenic rice and transient assay studies;

Figure 3 shows transverse sections from the mid-region of barley seeds (A-E) and steady state levels of the Ltp2 mRNA in different tissue fractions of developing barley endosperm;

Figure 4 shows the results for an *in situ* hybridization experiment;

Figure 5 is the result of a Southern blot experiment using DNA from transgenic rice plants;

Figure 6 shows the expression of a *GusA*-reporter gene driven by the Ltp2 gene promoter in the aleurone layer of developing transgenic rice seeds; and

Figure 7 shows the position of the myb and myc binding sites in the barley Ltp2 gene promoter.

A. METHODS

i. Plant material

Seeds of Barley (*Hordeum vulgare* cv. Bomi) were collected from plants grown in a phytotron as described before (Kvaale and Olsen, 1986). The plants were emasculated and pollinated by hand and isolated in order to ensure accurate determination of seed age.

ii. cDNA and genomic clones

The isolation and sequencing of the aleurone specific cDNA clone pBz11E (which is the same as Ltp2) was conducted as described by Jakobsen *et al.* (1989).

A barley, cv. Bomi genomic library was constructed by partial *Mbo*I digestion of total genomic DNA and subsequent ligation of the 10-20 kilo basepair (kb) size fraction with *Bam*HI digested lambda EMBL3 DNA (Clontech Labs, Palo Alto, Ca, USA). Out of a total of 2 x 10^6 plaques screened, using the Bz11E cDNA insert as a template for probe synthesis with a random labelling kit (Boehringer-Mannheim), four positive clones (gHv29-101, gHv38-201, gHv53-201 and gHv59-101) were identified after repeated rounds of plaque hybridization. DNA purified from these clones were restricted with several enzymes and characterized by Southern blot analysis.

The restriction maps of the four clones showed extensive overlap. One clone, gHv53-201, containing an insert of around 12 kb, was chosen for further analysis. A 6 kb *Pst*I fragment contained within the insert that hybridized to the cDNA probe was subcloned into Bluescript (Stratagene) giving the subclone BL53Ps17. A *Nhe*I restriction fragment of 0.7 kb covering the coding region of the Ltp2 gene was cloned into the *Xba*I site of M13mp18 and sequenced using the Sequenase protocol (USB) after isolation of ssDNA template using PCR amplification and magnetic beads (Dynabeads M280- Streptavidin, Dynal).

In order to characterize the 5' and 3' sequences of the Ltp2 gene, the following DNA fragments were generated by PCR amplification:

i) a 1.2 kb fragment covering the 5' end from a vector primer (KS) to the PLT11 primer located within the 5'end of the cDNA; and

ii) a 0.3 kb PCR product generated by amplification directed by the primers LTP13 and PLT15, of which the latter is based upon sequence information from the cDNA clone Bz14A, which is overlapping and identical with the Bz11E cDNA but contains some additional 30 base pairs after the polyadenylation site indicated at position 490 in Figure 1.

The sequences are:

KS:	5'	CGAGGTCGAC	GGTATCG		3'
PLT11:	5'	TACGG**T**GATC	TACTCGGCTA		3'
LTP13:	5'	ACGAAGCCGA	GCGGCGAGT		3'
PLT15:	5'	GGACTAAAAA	AAAAGTTGCA	ACACAAATTT C	3'.

The PLT11 sequence contains one base substitution (shown in bold and underlined) creating a *Bgl*II restriction site.

The 1.2 kb PCR product containing the 5' end was restricted with *Bgl*II which gave a 0.84 kb fragment with *Bam*HI compatible sticky ends that was subsequently cloned into the *Bam*HI site of pBluescript.

The 0.3 kb PCR product of the 3' end was treated with T4 DNA polymerase (Sambrook *et al.*, 1989) and subsequently cloned into the *Sma*I site of M13mp18.

The sequences of the PCR products were determined as described above.

iii. N rthern analyses

Total RNA was extracted from barley seed tissues of 10 DAP and older plant material essentially as described by Logemann *et al.* (1987), except that LiCl precipitation was used in place of ethanol precipitation. The RNA was denatured using formaldehyde and separated on 1.2% agarose gels as described by Selden (1987) and blotted onto GeneScreen (NEN) membranes using a Stratagene posiblotter apparatus according to supplier's instructions.

Hybridization was according to GeneScreen instruction manual (NEN) using radioactively labelled DNA strands complementary to the pBz11E cDNA insert generated with a random primed DNA labeling kit (Boehringer Mannheim).

iv. *In situ* hybridization

For *in vitro* transcription of antisense RNA, the plasmid pBz11E (Jakobsen *et al.*, 1989) was linearized with *Pst*I and transcribed with T7 RNA polymerase by using MAXIscript (Ambion) and [5,6-3H]-Uridine 5'-triphosphate (40-60 Ci mmol-1) (Amersham International) according to the specifications of the suppliers. The probe was hydrolyzed to fragments of about 100 bp as described by Somssich *et al.* (1988). Seed tissues were fixed in 1% glutaraldehyde, 100 mM sodium phosphate (pH 7.0) for 2 hours and embedded in Histoplast (Histolab, Göteborg, Sweden).

Sections of 10 μm were pretreated with pronase (Calbiochem) as described by (Schmelzer *et al.*, 1988) and hybridized with 25 ml of hybridization mix (200 ng probe ml-1, 50% formamide, 10% (w/v) dextran sulphate, 0.3 M NaCl, 10 mM Tris-HCl, 1 mM EDTA (pH 7), 0.02% polyvinyl- pyrrolidone, 0.2% Ficoll, 0.02% bovine serum albumin) for 15 hours at 50 °C.

Posthybridization was carried out according to Somssich *et al.* (1988) and autoradiography was done as described by Schmelzer *et al.* (1988), except that sections were exposed for 10 weeks.

v. Constructs for transient expression analysis

For the microprojectile bombardment experiments, the following constructs were used:

CONTROL A: pAct1f-*GUS* containing the rice *Actin 1* promoter fused to the *uidA* reporter gene encoding the *GUS* enzyme (McElroy *et al.*, 1990);

CONTROL B: pRT101-ex/s-int/s-LUC containing the *35S* CaMV promoter-*Sh1* first exon/intron fused to the firefly luciferase gene (Maas *et al.*, 1991); and

CONTROL C: pRT101C1 containing the *C1* cDNA downstream of the *35S* CaMV promoter (Paz-Ares *et al.*, 1987);

CONTROL D: pMF6Lc(R) containing the *Lc* cDNA corresponding to one *R* gene allele coupled to the *35S* CaMV promoter-*Adh1* intron (Ludwig *et al.*, 1989).

For the transient expression studies in barley aleurone the first intron of the maize *Sh1* gene carried on a 1.1 kb *Hinc*II fragment (Maas *et al.*, 1991) was inserted into the *Sma*I site of the promoter-reporter gene constructs according to the present invention. The Ltp2 gene promoter is contained on the 0.84 kb *Bgl*II fragment (sequence is presented in Figure 2) and was inserted into the *Bam*HI site of pBluescript. Thereafter the structural uidA gene encoding the beta-glucuronidase (*Gus*) enzyme was fused to the Ltp2 gene promoter.

The following conjugates according to the present invention were studied:

(i) Ltp2/GN: A Ltp2 gene promoter - *GUS* conjugate (same as conjugate in pLtp2/GN - see earlier);

(ii) Ltp2*Sh*1/GN: A Ltp2 gene promoter - Sh1 intron - *GUS* conjugate (same as conjugate in pLtp2*Sh*1/GN - see earlier).

Isolated plasmid DNA was used in the bombardment studies. The same conditions were used for the control conjugates and for the conjugates of the present invention.

For transient assay studies with rice protoplasts, the following conjugates according to the present invention were studied:

(i) Ltp2/GN: As above; and

(iii) Ltp2ΔBCIGN: A Ltp2 gene promoter {with a deletion spanning the myb and myc sites} - *GUS* conjugate (same as conjugate in pLtp2ΔBCIGN - see earlier).

vi. **Transformation of barley aleurone layers by particle bombardment**

Barley seeds were harvested at 25 DAP, surface sterilized in 1% sodium hypochlorite for 5 min and then washed 4 times in sterile distilled water. The maternal tissues were removed to expose the aleurone layer and the seed was then divided into two, longitudinally along the crease. The pieces of tissue were then placed, endosperm down, onto MS (Murashige & Skoog 1962) media with 10 g/l sucrose solidified with 10 g/l agar in plastic petri dishes (in two rows of 4 endosperm halves per dish).

Single bombardments were performed in a DuPont PDS 1000 device, with M-17 tungsten pellets (approx. 1 μm in diameter) coated with DNA as described by Gordon-Kamm *et al.* (1990) and using a 100 mm mesh 2 cm below the stopping plate. Equal amounts (25 μg per preparation) of the *GUS* (promoter-reporter gene) and LUC (internal standard) plasmids were mixed before adding the microprojectiles. One tenth of this amount, 2.5 μg, was used for the *Lc* and *C1* cDNA constructs. Bombarded tissue was incubated at 24°C for 3-4 days before extraction and measurement of *GUS* and LUC activities. Anthocyanin pigmentation could be observed in the bombarded aleurones directly without further treatment.
Histochemical staining for *GUS* expression was performed with X-Gluc (5-bromo,4-chloro,3-indolyl,ß-D,Glucuronic acid) as described by Jefferson (1987) at 37°C for 2 days. Extraction of cellular proteins for quantitative analysis was performed by grinding 4-8 half seeds in a mortar and pestle with 0.5 ml of extraction buffer (50 mM Na-phosphate pH, 1 mM DTT, pH 7.0).

After grinding, a further 0.5 ml was added and two 400 μl aliquots were taken. To one of these, 100 μl of 5 x Luciferase cell lysis buffer (Promega) was added and the sample vortexed before clearing by centrifugation at 10,000 rpm. A 20 μl aliquot was then measured for LUC activity in a scintillation counter (Tri-Carb 4000), using the luciferase assay system of Promega (E1500). To the other 400 μl aliquot, 100 μl of 5x lysis buffer (500 mM Na phosphate pH 7.0, 50 mM EDTA, 10 mM DTT, 0.5% Sarcosyl, 0.5% Triton X-100) was added, the mixture vortexed and cleared as above and assayed for *GUS* activity using 4-methylumbelliferone, ß-D,glucuronide as described by Jefferson (1987) modified to include 5% methanol in the reaction mixture (Kosugi *et al.*, 1990).

Production of 4-methylumbelliferone (MU) was measured after 1 and 4 h using a TKO 1000 Mini-Fluorimeter (Hoefer Scientific Instruments). In the analysis of promoter activities, the *GUS* readings (MU produced per hr) were standardized by division with the LUC value (photons produced over 30 s, beginning 60 s after mixing) from the same extract.

vii. Transient assay of rice protoplasts

In this experiment, the same type of protoplasts as used for stable transformation of rice plants was transiently transformed with constructs (i) and (iii) (see above) and then assayed for *GUS* activity.

viii. Rice transformation

Southern blot analysis of transgenic rice plants

Total genomic DNA was isolated from mature leaves, digested with *Xba* I and then transferred to a nylon membrane (Amersham). The coding region of the *GUS* gene was labelled and amplified with digoxigenin 11-dUTP by polymerase chain reaction and used for probing the Ltp2 - *GUS* gene. Hybridization and chemiluminesence signal detection were performed according to manufacturers specifications (Boeringen Mannheim).

B. RESULTS WITH REFERENCE TO THE FIGURES

i. **Figure 1** is a nucleotide sequence of the Ltp2 gene. A transcription start site has been assigned as +1. The TATA consensus sequence is boxed. Consensus myb and myc binding sites and the *SphI* element (Hattori *et al.*, 1992) found in the *C1* promoter sequence are shown in bold italics.

In the ORF (open reading frame), the nucleotides are shown in bold letters, starting with the first ATG codon and ending with the TAG stop codon. The single base substitution introduced at position +41 (A>T) creates a *BglII* restriction site which delimits the 3' end of the fragment covering the Ltp2 gene promoter. The positions of the 5' end and polyadenylation site of the corresponding cDNA, Bz11E (Jakobsen *et al.*, 1989), are shown by arrows. Two putative polyadenylation signals are underlined.

ii. **Figure 2a** is a nucleotide sequence for the Ltp2 gene promoter. Figure 2b is a nucleotide sequence for the Ltp2 gene promoter with an additional number of nucleotides for fusion to a *GUS* gene.

iii. **Figure 3** shows transverse sections from the mid-region of barley seeds (A-E) and steady state levels of the Ltp2 mRNA in different tissue fractions of developing barley endosperm (F).

Figure 3 can be analysed as follows:

> **(A)**: Ten DAP (days after pollination) endosperm isolated from the surrounding maternal tissues by manual extrusion. For maternal tissues, see (C). The extruded endosperm consists of the central starchy endosperm cells, a group of modified aleurone cells over the crease area (arrow) and one layer of highly vacuolated peripheral aleurone cells (arrowhead).
>
> **(B)**: Enlargement showing vacuolated peripheral aleurone cells (AC) and starchy endosperm cells (SE) in area of (A) marked with arrowhead.

(C): Pericarp of 10 DAP seed after extrusion of the endosperm with the nucellus epidermal layer (NE) facing the endosperm cavity, which contained the endosperm in (A) before extrusion.

(D): Extruded 15 DAP endosperm with central starchy endosperm cells and modified aleurone cells (arrow), but without peripheral aleurone cells.

(E): 15 DAP pericarp with adhering aleurone layer after extrusion of the endosperm (in D).

(F): Northern blot showing the steady state level of Ltp2 mRNA in the extruded endosperm fraction (e) and the pericarp fraction (p) in the interval from 10 to 13 DAP. For this blot, 10 μg of total RNA was loaded in each lane. The gel was blotted and hybridized with randomly primed Ltp2 cDNA.

iv. **Figure 4** shows the results for an *in situ* hybridization of ^{3}H-labelled Ltp2 antisense probe to transverse sections of barley endosperm (A and B) and transient gene expression analysis of different promoter-reporter gene constructs in developing barley aleurone layers after particle bombardment (C, D and E). Figure 4 can be analysed as follows:

(A): Dark field microphotograph of 13 DAP endosperm showing hybridization of the Ltp2 probe to the peripheral aleurone cells (AL) ventrally and laterally, but not to aleurone cells on the dorsal side of the grain (DS), nor to the modified aleurone cells over the crease area (MA).

(B): Magnification of peripheral endosperm (frame in A) showing gradient of *in situ* hybridization signal towards the dorsal side of the seed containing undifferentiated aleurone cells.

(C): Colourless barley aleurone layer co-bombarded with the *35S-C1* and *35S-Lc* cDNA constructs. Single aleurone cells synthesizing anthocyanin pigment appear as red spots.

(D): Exposed aleurone layer of 25 DAP barley seeds bombarded with the Ltp2/*Sh1* int/*GUS* construct. The transfected seed was stained for detection of *GUS* activity.

(E): Exposed aleurone layer of barley seed of the same stage bombarded with *pAct1f-GUS* construct and histochemically stained as in (D).

(V): Ventral crease area.

v. **Figure 5** is the result of a Southern blot experiment of DNA from transgenic rice plants harbouring the Ltp2-*GUS* construct. Figure 5 can be analysed as follows:

Lane P: plasmid Ltp2-*GUS*.

Lane C: untransformed control plants.

Lane 1: transgenic line 3-15.

Lane 2: transgenic line 4-13.

Lane 3: transgenic line 2-6.

Lane 4: transgenic line 4-6.

vi. **Figure 6** shows the expression of a *GUS*-reporter gene driven by the Ltp2-wildtype promoter in the aleurone layer of developing transgenic rice seeds. Figure 6 can be analysed as follows:

(A): Longitudinal section of 20 DAP seed showing *GUS* staining exclusively in the aleurone layer (AL), but not in the embryo, starchy endosperm (SE) or in the maternal tissue (M).

(B): Transverse section from the mid-region of 20 DAP seed.

(C): Enlargement of dorsal side of seed shown in (A).

(D): Non-transgenic control seed, same age as in (A).

(E): A 5 day-old seedling transformed with the Ltp2 - *GUS* gene.

(F): A 5 day-old seedling transformed with the CaMV35S - *GUS* gene. (Terada and Shimamoto 1990). Arrows indicate regions of *GUS* expression. Bars in (A,B and D) are 10 mm and in (C) 2.5 mm.

vii. **Figure 7** shows the position of the myb and myc sites in the barley Ltp2 gene promoter. The distance from the 3' end of the myc site to the TATA box is given in nucleotides. The following nucleotides from and between the myb and myc sites were deleted to form the conjugate containing the deletion in the Ltp2 gene promoter gene:

CAACTACCATCGGCGAACGACCCAGC.

C. CONCLUSIONS

1. *The barley Ltp2 gene encodes a protein homologous to the 7 kDa wheat lipid transfer protein*

Using the Bz11E cDNA (Jakobsen *et al.*, 1989) as a probe, the corresponding barley cv. Bomi genomic clone was isolated. The sequences of the genomic clone and that of the Bz11E cDNA are identical in overlapping regions and no intervening sequences were detected (Figure 1) accordingly this gene is Ltp2. The ATG codon initiating the longest open reading frame (ORF) in the Ltp2 sequence is located 64 bp downstream of the putative transcriptional start site at nucleotide number 1 (Figure 1). The ORF contains eight potential translation start codons between nucleotides 64 and 127. Two polyadenylation signals, which conform to the plant consensus sequence (Joshi, 1987) are found in the 3' end of the genomic sequence. In the Bz11E cDNA the polyA tail extends after the G at position 491 (Figure 1 and Figure 2).

2. *The Ltp2 transcript can be a molecular marker for peripheral aleurone cell differentiation*

In the developing seed at approximately 8 days after pollination (DAP), aleurone cell differentiation is initiated over the ventral crease area resulting in the formation of the modified aleurone cells (Figure 3A and Bosnes *et al.*, 1992). Shortly after, at 9 DAP, the second type of aleurone cells, characterized by their extensive vacuolation (Figure 3B), appears in the peripheral endosperm close to the crease area, spreading first laterally and then to the dorsal side of the seed (see Figure 3A). At this stage, when whole de-embryonated seeds are squeezed, the extruded endosperm consists of the starchy endosperm, the peripheral and the modified aleurone cells (Figure 3 A-C). This is in contrast to later developmental stages, where the extruded endosperm consists only of the starchy endosperm and the modified aleurone cells (Figure 3D). The reason for this is that the cells of the aleurone layer adhere to the maternal pericarp (Figure 3E). Aleurone cell formation is completed at 21 DAP, when cell division stops (Kvaale and Olsen, 1986). Using the Ltp2 probe on Northern blots with total RNA, the signal obtained gradually becomes stronger in the pericarp, compared to the extruded endosperm, confirming the relocation of the aleurone cells from the endosperm fraction to the pericarp fraction in the interval between 10 and 13 DAP (Figure 3F).

From the experimental results presented in Figure 3 it is concluded that the Ltp2 transcript is a potential marker for aleurone cell differentiation. To corroborate the usefulness of the Ltp2 transcript as a molecular marker for aleurone cell differentiation, *in situ* analysis was carried out on transverse sections of 13 DAP seeds. The rationale for using seeds from this stage was the earlier observed gradual differentiation of the peripheral aleurone cells, starting near the crease area and spreading to the dorsal side (Bosnes *et al.*, 1992).

Using ^{3}H-labelled antisense transcript as probe, a positive signal is clearly visible in the peripheral aleurone cells in the ventral part adjacent to the crease area as well as laterally up towards the dorsal side of the grain (Figure 4A). However, no signal is present in the dorsal region of the seed, nor over the modified aleurone cells.

Focusing on the most dorsal aleurone cells showing a positive signal in the *in situ* analysis (Figure 4B), the morphology of these cells corresponds to that of the highly vacuolated peripheral aleurone cells in 10 DAP endosperm (Figure 3B).

The Ltp2 transcript therefore represents a highly tissue specific molecular marker for aleurone cell differentiation.

3. *The Ltp2 gene promoter is transiently expressed in developing barley aleurone cells after particle bombardment*

The Ltp2 gene promoter contained on a 842 bp *BglII* restriction fragment (from nt -807 to nt +35 in Fig.1) was fused to the *GUS*-reporter gene and introduced into the exposed aleurone layers of 25 DAP whole barley seeds by the biolistic method. In the first set of experiments, Ltp2 gene promoter activity was assayed visually after histochemical staining with X-Gluc. Due to the large variation between individual experiments with the biolistic method, plasmid DNA containing the *Lc* and *C1* cDNAs from maize under the control of the 35S CaMV promoter was co-bombarded with the Ltp2 construct to monitor shooting efficacy. In combination, but not individually, the latter two cDNAs give high numbers of red anthocyanin spots in the barley aleurone cells without any treatment after 1 to 2 days of incubation of the seeds on solid nutrient medium (Figure 4C). Compared to the number of red spots, the Ltp2-*GUS* construct consistently gave very few spots after histochemical staining in co-bombardment experiments.

Based on previous reports that insertion of introns in promoter construct enhance the level of transient expression (Maas *et al.*, 1991) without interfering with the tissue specificity of the promoters, the intron from the maize *Shrunken-1* gene was inserted into the Ltp2-*GUS* construct after the promoter. Using this construct the number of spots in immature aleurone layers increased (Figure 5D). Still, however, compared to aleurone layers bombarded with the *pAct1f-GUS* construct (McElroy *et al.*, 1990), which contains the promoter of the constitutively expressed *Actin1* gene from rice (Figure 4D), both the number and the size of the spots obtained with the Ltp2-*GUS* construct is dramatically smaller (Figure 4E).

In order to quantify Ltp2 gene promoter activity in particle bombardment experiments, another control plasmid containing the *LUC* gene under the control of the 35S-promoter was co-bombarded with the Ltp2-*GUS* constructs. In this way, after particle bombardment and incubation on tissue culture medium, protein was extracted from the seeds in a buffer that allowed measurement of both LUC and *GUS* activity (for details, see Materials and Methods section). In such experiments, calculating *GUS* expression standardized on the base of the LUC-activity, the Ltp2-*GUS* activity was not significantly higher than background, corresponding approximately to 1.5% of the Actin1f promoter activity in parallel experiments.

For the Ltp2-*Sh1* intron-*GUS* construct, however, the activity was significantly higher than background, corresponding to 5% of that of the *Actin1* promoter. Blue spots from the activity of the Ltp2-promoter were never observed in other tissues than the aleurone layer of developing seeds. From these experiments it is concluded that the -807 bp promoter of the Ltp2 gene is capable of directing transient gene expression in a fashion similar to that of the endogenous Ltp2 gene, i.e., in the cells of the aleurone layer of immature barley seeds.

4. The Ltp2 gene promoter directs aleurone specific expression of the GUS-reporter gene in transgenic rice seeds

The gene was transformed into rice by electroporation of embryogenetic protoplasts following the teachings of Shimamoto et al. 1989.

Four fertile transgenic rice plants were obtained and integration of the Ltp2-*GUS* gene was examined by Southern blot analysis. The results demonstrated that a 2.9 b fragment containing the Ltp2-*GUS* gene is integrated in all the transgenic lines. Histochemical *GUS* analysis was carried out with developing rice seeds of 20 DAP and 5 day old seedings derived from transgenic seeds (Figure 6). In developing seeds the *GUS* expression is strictly limited to the aleurone layer, with no staining observed in the maternal tissues, starchy endosperm or in the embryo of the transgenic seeds (Figure 6 A-C), nor in untransformed control seeds (Figure 6 D). No *GUS* staining was observed in leaves or roots of seedlings transformed with the Ltp2 - *GUS* gene (Figure 6 E).

In contrast, seedlings transformed with the CaMV35S - *GUS* gene *GUS* expression is detected in the coleptile, shoots and roots (see Figure 6 F; Terada and Shimamoto 1990).

These results clearly demonstrate the aleurone-specific expression of the Ltp2 - *GUS* gene in transgenic rice plants.

5. The Ltp2 gene promoter contains sequence elements implicated in the transcriptional control of cereal endosperm specific genes

The studies with the deletion spanning the myb and myc sites in the Ltp2 gene promoter showed that levels of expression were about 10% of that of the wild-type gene promoter. These studies indicated that both the myb and myc sites are important for expression.

In addition, the Ltp2 gene promoter may even contain another sequence element that has been implicated in regulation of gene expression in maize aleurone cells, namely the octanucleotide CATGCATG (Figure 1). This sequence, previously referred to as the *Sph*I element, has been shown to mediate the transcriptional activation of maize *C1* by interaction with VP1 (Hattori *et al.*, 1992). As in the maize *C1* promoter (Paz-Ares *et al.*, 1987), the putative *Sph*I element of the barley Ltp2 gene promoter is located between the TATA-box and the myb binding site.

In addition, the Ltp2 gene promoter may contain two further sites that could play an important role in transcription. The first site is an "AL" site and has the sequence

CATGGAAA

This AL sequence ends at position -366 in the sequence shown in Figure 1.

The second site is a "DS" site that has a high degree of similarity or identity with the binding site for 5' transcriptional factors from other eucaryotic organisms. This DS site, which a dyad-symmetry, has the sequence

TCGTCACCGACGA

This DS sequence ends at position -121 in the sequence shown in Figure 1.

D. DISCUSSION

The above examples of the present invention concern the barley gene Ltp2, which encodes an aleurone specific 7 kDa nsLTP.

The identification of the Ltp2 protein as a nsLTP is based on the high identity (78%) between the predicted Ltp2 amino acid sequence and the 7 kDa protein isolated from wheat seeds using the *in vitro* lipid transfer assay (Monnet, 1990). The high degree of sequence identity between the two barley aleurone Ltp gene products and the homologous proteins and transcripts from wheat seeds strongly suggests that the aleurone layer of these two cereals contain two related classes of nsLTPs with molecular masses of 10 and 7 kDa, respectively.

While the sequence identity is more than 70% within the two classes, it is only around 20% between them. However, several conserved features are apparent in the cereal seed nsLtps, including similar N-terminal signal peptides, an internal stretch of 20 amino acids with 60% similarity, and 8 cysteine residues that are believed to be important for the activity of plant Ltps (Tchang *et al.*, 1988). Studies also showed that the Ltp2 gene lacks an intron. Hybridization experiments using Ltp2 probes to barley genomic Southern blots indicate that the barley haploid genome contains only one copy of each gene (Jakobsen *et al.*, 1989; Skriver *et al.*, 1992).

According to a suggestion by Sterk *et al.* (1991) plant nsLTPs may be involved in the extracellular transport of cutin or other lipid monomers from the endoplasmic reticulum to the site of synthesis of extracellular matrix components, such as the cuticle. Therefore, one possible role for the aleurone specific nsLTPs in barley and wheat could be in the formation of the earlier described amorphous layer on the outside of the aleurone cells in wheat seeds (Evers and Reed, 1988). The function of this layer is unknown, but it may be involved in the regulation of the osmotic pressure in the endosperm during seed development and germination. If this holds true, the absence of the Ltp2 transcript in the modified aleurone cells in the ventral crease area is functionally significant, since an impermeable layer on the outside of these cells would prevent the influx of soluble synthates from the vegetative plant parts.

Of the nine cDNAs isolated in the differential screening experiment design to identify clones representing transcripts differentially expressed in the aleurone layer of developing barley seeds, Ltp2 hybridizes to transcripts present exclusively in the aleurone layer. Thus, the Ltp2 gene represents a suitable gene for the search for promoter sequences responsible for the control of gene transcription in the aleurone layer.

Due to the lack of a routine protocol for stable barley transformation, demonstration of Ltp-promoter specificity in barley has to rely on transient assays using the particle bombardment method. Using this method, it was demonstrated that the -807 bp Ltp2 gene promoter carried on the *Bgl*II restriction fragment is capable of driving the expression of the *GUS* reporter gene in immature barley aleurone layers. From this it is concluded that the promoter fragment carries sequences that are responsible for barley aleurone specific gene transcription.

The Ltp2 gene promoter can be weaker than constitutive cereal promoters like that of the *Actin1f* gene - even after the introduction of the *Sh1*-intron (see Maas *et al.* (1991) and their work on tobacco protoplasts) into the Ltp2-*GUS* construct which increases the expression levels by around three-fold. However, this lower expression does not result in any damage to the developing seedling - unlike the constitutive cereal promoters. Moreover, and again unlike the constitutive cereal promoters, the Ltp2 gene promoter directs desirable tissue and stage specifc expression.

As demonstrated by the histochemical assays shown in Figure 6, the Ltp2 *Bgl*II promoter fragment shows the same aleurone specific expression in developing rice seeds as in barley.

Thus, the conclusion from the transient assays in barley that this promoter fragment contains sequences responsible for aleurone specific gene transcription is confirmed. Furthermore, the data from rice provide support to the view that the molecular mechanisms underlying aleurone specific gene transcription in developing seeds are conserved among the cereal species.

E. SUMMATION

The Examples describe the isolation of the promoter for the barley gene Ltp2, which encodes a novel class of cereal 7 kDa nsLTPs. The gene was isolated by the use of a cDNA from a differential screening experiment in which the positive probe was constructed from aleurone cell poly (A) rich RNA, and the negative probe from the starchy endosperm of immature seeds.

In situ hybridization analysis demonstrates that the Ltp2 transcript is expressed exclusively in the aleurone layer from the beginning of the differentiation stage and half way into the maturation stage. Similar to previously identified 10 kDa plant nsLTPs, the Ltp2 protein contains the eight conserved cysteine residues.

The results indicate that the Ltp2 protein is involved in the synthesis of a lipid layer covering the outside of the cereal aleurone cells.

Using particle bombardments it was shown that the -807 bp Ltp2 gene promoter fused to the *GUS*-reporter gene is active in the aleurone layer of developing barley seeds, giving 5% of the activity of the strong constitutive actin1f-promoter from rice. Transformed into rice, the barley Ltp2-promoter directs strong expression of the *GUS*-reporter gene exclusively in the aleurone layer of developing rice seeds. Analysis of the Ltp2 gene promoter reveals the presence of sequence motives implicated in endosperm specific gene expression in maize, i.e. the myb and myc protein binding sites. In short, the Ltp2 gene promoter represents a valuable tool for the expression of GOIs in the aleurone layers of cereal seeds.

Other modifications of the present invention will be apparent to those skilled in the art without departing from the scope of the invention.

REFERENCES

Baker, R.E, Gabrielsen, O.S., and Hall, B.D. (1986) Effects of tRNATyr point mutations on the binding of the yeast RNA polymerase III transcription factor C. *J. Biol. Chem.* **261**, 5275 - 5282.

Barkardottir, R.B., Jensen, B.F., Kreiberg, J.D., Nielsen, P.S. and Gausing, K. (1987) Expression of selected nuclear genes during leaf development in barley. *Dev. Genet.* **8**, 495 - 511.

Bosnes, M., Weideman, F. and Olsen, O.-A. (1992) Endosperm differentiation in barley wild-type and sex mutants. *Plant J.* **2**, 661-674.

Dellaporta, S.L., Greenblatt, I., Kermicle, J.L., Hick, J.B. and Wessler, S. (1988) Molecular cloning of the *R-nj* allelel by transposon tagging with *Ac*. In *Chromosome structure and function: Impact of new concepts*, 18th Stadler Genetics Symposium (Gustafson, J.P. and Appels, R., eds.), New York: Plenum Press, pp. 263-282.

Dieryck, W., Gautier, M.-F., Lullien, V. and Joudrier, P. (1992) Nucleotide sequence of a cDNA encoding a lipid transfer protein from wheat (*Triticum durum* Desf.). *Plant Mol. Biol.* **19**, 707-709.

Dooner, H.K. (1983) Coordinate genetic regulation of flavonoid biosynthetic enzymes in maize. *Mol. Gen. Genet.* **198**, 136-14.

Dooner, H.K. (1985) *Viviparous -1* mutation in maize conditions pleiotropic enzyme deficiencies in the aleurone. *Plant Physiol.* **77**, 486-488.

Evers, A.D. and Reed, M. (1988) Some novel observations by scanning electron microscopy on the seed coat and nucellus of the mature wheat grain. *Cereal Chem.* **65**, 81-85.

Fincher, G.B. (1989) Molecular and cellular biology associated with endosperm mobilization in germinating cereal grains. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 305-346.

Fleming, A.J, Mandel, T., Hofmann, S., Sterk, P., de Vries, S.C., and Kuhlemeier, C. (1992) Expression pattern of a tobacco lipid transfer protein gene within the shoot apex. *Plant J.* **2**, 855-862.

Franken, P., Niesbach-Kløsgen, U., Weydeman, U., Marechal-Drouard, L., Saedler,H., and Wienand, U. (1991). The duplicated chalcone synthase genes *C2* and *Whp* (*white pollen*) of *Zea mays* are independently regulated; evidence for translational control of *Whp* expression by the anthocyanin intensifying gene . *EMBO J.* **10**, 2605-2612.

Friedman, W.E. (1992) Evidence of a pre-angiosperm origin of endosperm: implication for the evolution of flowering plants. *Science* **255**, 336-339.

Gabrielsen, O.S., Sentenac, A. and Fromageot, P. (1991) Specific DNA binding by c-MYB: Evidence for a double helix-turn-helix-related motif. *Science*, **253** 1140-1143.

Goff, S.A., Cone, K.C. and Chandler, V.L. (1992) Fuctional analysis of the transcriptional activator encoded by the maize *B* gene: evidence for a direct functional interaction between to classes of regulatory proteins. *Genes Dev.* **6**, 864-875.

Gordon-Kamm, W.J., Spencer, T.M., Mangano, M.L., Adams, T.R., Daines, R.J., Start, W.G., O'Brien, J.V., Chambers, S.A., Adams, W.R., Willetts, N.G., Rice, T.B., Mackey, C.J., Krueger, R.W., Kausch, A.P. and Lemaux, P.G. (1990) Transformation of maize cells and regeneration of fertile transgenic plants. *Plant Cell* **2**, 603-618.

Hammond-Kosack, M.C.U., Holdsworth, M.J. and Bevan, M. (1993) *In vivo* foot-printing of a low molecular weigth glutenin gene (LMWG-1D1) in wheat endosperm. *EMBO J.* **12**, 545-554.

Hattori, T., Vasil, V., Rosenkrans, L., Hannah, L.C., McCarty, D.L. and Vasil, I.K. (1992) The *Viviparous-1* gene and abscisic acid activate the *C1* regulatory gene for anthocyanin biosynthesis during seed maturation in maize. *Genes Dev.* **6**, 609-618.

Jacobsen, J.V., Knox, R.B. and Pyliotis, N.A. (1971) The structure and composition of aleurone grains in the barley aleurone layer. *Planta* **101**, 189-209.

Jakobsen, K., Klemsdal, S., Aalen, R., Bosnes, M., Alexander, D. and Olsen, O.-A. (1989) Barley aleurone cell development: molecular cloning of aleurone-specific cDNAs from immature grains. *Plant Mol. Biol.* **12**, 285-293.

Jefferson, R.A. (1987) Assaying chimeric genes in plants: the *GUS* gene fusion system. *Plant Mol. Biol. Rep.* **5**, 387-405.

Joshi, C.P. (1987) Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. *Nucleic Acid Res.* **15**, 9627-9640.

Kalla, R., Lonneborg, A., Linnestad, C,. Potter, R., Aalen, R., Nielsen, P.S. and Olsen, O.-A. (1993) Characterisation of promoter elements of aleurone specific genes from barley. In *Pre-harvest sprouting in cereals 1992*. Eds. M.K.Walker-Simmons and J.J.Ried. The American Association of Cereal Chemists, Inc. ISBN 0-913250-81-3. pp 236-245.

Karrer, E.J., Litts, J.C and Rodriguez, R.L. (1991) Differential expression of a-amylase genes in germinating rice and barley seeds. *Plant Mol. Biol.* **16**, 797-805.

Kader, J.-C., Julienne, M. and Vergnolle, C. (1984) Purification and characterisation of a spinach-leaf protein capable of transferring phospholipids from liposomes to mitochondria or chloroplasts. *Eur. J. Biochem.* **139**, 411-416.

Koltunow, A.M., Truettner, J., C x, K.H., Wallroth, M. and Goldberg, R.B. (1990) Different temporal and spatial expression patterns occur during anther development. *Plant Cell* 2, 1201-1224.

Kosugi, S., Ohashi, Y., Nakajima, K. and Arai, Y. (1990) An improved assay for b-glucuronidase in transformed cells: methanol almost completly suppresses a putative endogenous b-glucuronidase activity. *Plant Sci.* **70**, 133-140.

Kvaale, A. and Olsen, O.-A. (1986) Rates of cell division in developing barley endosperms. *Ann. Bot.* **57**, 829-833.

Kyozuka, J. and Shimamoto, K. (1991) Transformation and regeneration of rice protoplasts. In *Plant Tissue Culture Manual* B1 (Lindsey, K., ed.) Dordrecht: Kluwer Academic Publishers, pp. 1-6.

Kyozuka,J., Fujimoto, H., Izawa,T. and Shimamoto, K. (1991). Anaerobic induction and tissue-specific expression of maize *Adh*1 promoter in transgenic rice plants and their progeny. *Mol.Gen.Genet.* **228**, 40-48.

Lanahan, M.B., Ho, T-H.D., Rogers, S. and Rogers, J. (1992) A Gibberellin response complex in cereal a-amylase gene promoters. *Plant Cell* **4**, 203-211.

Lea, R., Tommerup, H., Svendsen, I. and Mundy, J. (1991) Biochemical and molecular characterization of three barley seed proteins with antifungal properties. *J. Biol. Chem.* **266**, 1564-73.

Linnestad, C., Lönneborg, A., Kalla, R. and Olsen, O.-A. (1991) The promoter of a lipid tranfer protein gene expressed in barley aleurone cells contains similar Myb and Myc recognition sites as the maize *Bz-McC* allele. *Plant Physiol.* **97**, 841-843.

Logemann, J., Schell, J. and Willmitzer, L. (1987) Improved method for isolation of RNA from plant tissues. *Anal. Biochem.* **163**, 16-20.

Ludwig, S.R., Habera, L.F., Dellaporta, S.L. and Wessler, S.R. (1989) *Lc*, a member of the maize *R* gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the *myc*-homology region. *Proc. Natl. Acad. Sci. USA* **86**, 7092-7096.

Maas, C., Laufs, J., Grant, S., Korfhage, C. and Werr, W. (1991) The combination of a novel stimulatory element in the first exon of the maize *shrunken-1* gene with the following intron1 enhances reporter gene expression up to 1000 fold. *Plant. Mol. Biol.* **16**, 199-207.

Madrid, S.M. (1991) The barley lipid transfer protein is targeted into the lumen of the endoplasmic reticulum. *Plant Physiol.Biochem.* **29**, 695-703.

Marocco, A., Wissenbach, M., Becker, D., Paz-Ares, J., Saedler, H. and Salamini, F. (1989) Multiple genes are transcribed in *Hordeum vulgare* and *Zea mays* that carry the DNA binding domain of the myb oncoproteins. *Mol. Gen. Genet.* **216**, 183-187.

McCarty, D.R., Carson, C.B., Stinard, P.S. and Robertson, D.S. (1989) Molecular analysis of *Viviparous-1*: An abscisic acid-insensitive mutant of maize. *Plant Cell* **1**, 523-532.

McCarty, D.R., Hattori, T., Carson, C.B., Vasil, V., Lazar, M. and Vasil, I.K. (1991) The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. *Cell* **66**, 895-905.

McClintock, B. (1978) Development of the maize endosperm as revealed by clones. In *The clonal basis of development* (Subtelny, S. and Sussex, I.M., eds.), New York: Academic Press, pp.217-237.

McElroy, D., Zhang, W., Cao, J. and Wu, R. (1990) Isolation of an efficient actin promoter for use in rice transformation. *Plant Cell* **2**, 163-171.

Monnet, F.-P. (1990) Ph.D thesis. Universite des Sciences et Techniques du Languedoc, Montpellier, France, pp. 121.

Mundy, J. and Rogers, J. (1986) Selective expression of a probable amylase/protease inhibitor in barley aleurone cells: comparison to the barley amylase/subtilisin inhibitor. *Planta* **169**, 51-62.

Murashige, T., and Skoog, F. (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473-497.

Olsen, O.-A., Potter, R.H. and Kalla, R. (1992) Histo-differentiation and molecular biology of developing cereal endosperm. *Seed Sci. Res.* **2**, 117-131.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A. and Saedler, H. (1987) The regulatory *c1* locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J.* **6**, 3553-3558.

Qu, R., Wang, S.-M., Lin, Y.-H., Vance, V.B. and Huang, A.H.C. (1986) Characteristics and biosynthesis of membrane proteins of lipid bodies in the scuttela of maize. *Biochem J.* **235**, 57-65.

Rohde, W., Dörr, S., Salamini, F. and Becker, B. (1991) Structure of a chalcone synthase gene from *Hordeum vulgare*. *Plant Mol. Biol.* **16**, 1103-1106.

Roth, B.A., Goff, S.A., Klein, T.M. and Fromm, M.E. (1991) *C1*- and *R*-dependent expression of the maize *Bz1* gene requires sequences with homologies to mammalian myb and myc binding sites. *Plant Cell* **3**, 317-325.

Sambrook, L., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning, a Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.

Schmelzer, E., Jahnen, W. and Hahlbrock, K. (1988) *In situ* localization of light-induced chalcone synthetase mRNA, chalcone synthetase, and flavonoid end products in epidermal cells of parsley leaves. *Proc. Natl. Acad. Sci. U.S.A* **85**, 2989-2993.

Selden, R.F. (1987) Analysis of RNA by Northern Hybridization. In *Current Protocols in Molecular Biology* (Ausubel, F.M. *et al.*, eds.). New York: Green Publishing Associates and Wiley-Interscience, pp. 4.9.1-4.9.8.

Shimamoto, K., Terada, R., Izawa, T. and Fujimoto, H. (1989) Fertile transgenic rice plants regenerated from transformed protoplasts. *Nature* **338**, 274-276.

Skriver, K., Leah, R., Müller-Uri, F., Olsen, F.-L. and Mundy, J. (1992) Structure and expression.

Slakeski, N. and Fincher, G.B. (1992) Developmental regulation of (1-3,1-4)-b-glucanase gene expression in barley. *Plant Physiol.* **99**, 1226-1231.

Smith, L.M., Handley, J., Li, Y., Martin, H., Donovan, L. and Bowles, D.J. (1992) Temporal and spatial regulation of a novel gene in barley embryos. *Plant Mol. Biol.* **20**, 255-266.

Somssich, I.E., Schmelzer, E., Kawalleck, P. and Hahlbrock, K. (1988) Gene structure and *in situ* transcript localization of the pathogenesis-related protein 1 in parsley. *Mol. Gen. Genet.* **213**, 93-98.

Sossountzov, L., Riuz-Avila, L., Vignois, F., Jolliot, A., Arondel, V., Tchang, F., Grosbois, M., Guerbette, F., Miginiac, E., Delsney, M., Puigdomenech, P. and Kader, J.-C. (1991) Spatial and temporal expression pattern of a maize lipid transfer protein gene. *Plant Cell* **3**, 923-933.

Sterk, P., Booij, H., Schellekens, G.A., Van Kammen, A. and De Vries, S.C. (1991) Cell-specific expression of the carrot EP2 lipid transfer protein gene. *Plant Cell* 3, 907-921.

Terada, R. and Shimamoto,K. (1990) Expression of CaMV35S-*GUS* gene in transgenic rice plants. *Mol.Gen.Genet.* **220**, 389-392.

Terada,R., Nakayama,T., Iwabuchi,M., Shimamoto, K. (1993) A wheat histone H3 promoter confers cell division-dependent and -independent expression of the *GUS*A gene in transgenic rice plants. *Plant J* 3: 241-252.

Tchang, F., This, P., Stiefel, V., Arondel, V., Morch, M.D., Pages, M., Puigdomenech, P., Grellet, F., Delsney, M., Bouillon, P., Huet, J.C., Guerbette, F., Beauvais-Cante, F., Duranton, H., Pernollet, J.C. and Kader, J.-C. (1988) Phospholipid transfer protein: Full-length cDNA and amino acid sequence in maize. *J. Biol. Chem.* **263**, 16849-16855.

Thoma,S., Kaneko,Y. and S mmerville,C. (1993) A non-specificlipid transfer protein from Arabidopsis is a cell wall protein. *The Plant Journal* **3(3)**, 427-436.

von Heijne, G. (1988) Transcending the impenetrable: How proteins come to term with the membranes. *Biochim.Biophys. Acta.* **947**, 307-333.

Watanabe, S. and Yamada, M. (1986) Purification and characterization of a non-specific lipid transfer protein from germinated castor bean endosperms which transfers phospholipids and galactolipids. *Biochim. Biophys. Acta.* **876**, 116-123.

Weston, K. (1992) Extension of the DNA binding consensus of the chicken c-Myb and v-Myb proteins. *Nucleic Acids Res.* **20**, 3043-3049.

SEQUENCE LISTING

(1) GENERAL INFORMATION

NAME OF APPLICANTS: O.-A. OLSEN AND R. KALLA
BUSINESS ADDRESS: PLANT MOLECULAR BIOLOGY LABORATORY
DEPARTMENT OF BIOTECHNICAL SCIENCES
AGRICULTURAL UNIVERSITY OF NORWAY AND
AGRICULTURAL BIOTECHNOLOGY PROGRAM NRC
NORWAY N-1432

TITLE OF INVENTION: PROMOTER

(2) INFORMATION FOR SEQUENCE I.D. 1

SEQUENCE TYPE: NUCLEIC ACID
MOLECULE TYPE: DNA (GENOMIC)
ORIGINAL SOURCE: BARLEY
SEQUENCE LENGTH: -807
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
SEQUENCE:

-807		GATCTCG	ATGTGTAGTC	TACGAGAAGG
-780	GTTAACCGTC	TCTTCGTGAG	AATAACCGTG	GCCTAAAAAT
	AAGCCGATGA	GGATAAATAA	AATGTGGTGG	TACAGTACTT
	CAAGAGGTTT	ACTCATCAAG	AGGATGCTTT	TCCGATGAGC
-660	TCTAGTAGTA	CATCGGACCT	CACATACCTC	CATTGTGGTG
	AAATATTTTG	TGCTCATTTA	GTGATGGGTA	AATTTTGTTT
	ATGTCACTCT	AGGTTTTGAC	ATTTCAGTTT	TGCCACTCTT
-540	AGGTTTTGAC	AAATAATTTC	CATTCCGCGG	CAAAAGCAAA
	ACAATTTTAT	TTTACTTTTA	CCACTCTTAG	CTTTCACAAT
	GTATCACAAA	TGCCACTCTA	GAAATTCTGT	TTATGCCACA
-420	GAATGTGAAA	AAAAACACTC	ACTTATTTGA	AGCCAAGGTG
	TTCATGGCAT	GGAAATGTGA	CATAAAGTAA	CGTTCGTGTA
	TAAGAAAAAA	TTGTACTCCT	CGTAACAAGA	GACGGAAACA
-300	TCATGAGACA	ATCGCGTTTG	GAAGGCTTTG	CATCACCTTT
	GGATGATGCG	CATGAATGGA	GTCGTCTGCT	TGCTAGCCTT
	CGCCTACCGC	CCACTGAGTC	CGGGCGGCAA	CTACCATCGG
-180	CGAACGACCC	AGCTGACCTC	TACCGACCGG	ACTTGAATGC
	GCTACCTTCG	TCAGCGACGA	TGGCCGCGTA	CGCTGGCGAC
	GTGCCCCCGC	ATGCATGGCG	GCACATGGCG	AGCTCAGACC
-60	GTGCGTGGCT	GGCTACAAAT	ACGTACCCCG	TGAGTGCCCT
	AGCTAGAAAC	TTACACCTGC		

NOTE: ABOVE SEQUENCE IS A RETYPED VERSION OF FIGURE 2A WHICH IS TO BE TAKEN AS THE CORRECT SEQUENCE

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 11, line 6

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution

The National Collections of Industrial and Marine Bacteria Limited (NCIMB)

Address of depositary institution *(including postal code and country)*

23 St. Machar Drive
Aberdeen
Scotland
AB2 1RY
United Kingdom

Date of deposit	Accession Number
22 November 1993	NCIMB 40598

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet ☐

In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international applicati n	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 11, line 12

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution

The National Collections of Industrial and Marine Bacteria Limited (NCIMB)

Address of depositary institution *(including postal code and country)*

23 St. Machar Drive
Aberdeen
Scotland
AB2 1RY
United Kingdom

Date of deposit	Accession Number
22 November 1993	NCIMB 40599

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 11 **, line** 20

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution

The National Collections of Industrial and Marine Bacteria Limited (NCIMB)

Address of depositary institution *(including postal code and country)*

23 St. Machar Drive
Aberdeen
Scotland
AB2 1RY
United Kingdom

Date of deposit	Accession Number
22 November 1993	NCIMB 40601

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet ☐

In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 11, line 28

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution

The National Collections of Industrial and Marine Bacteria Limited (NCIMB)

Address of depositary institution *(including postal code and country)*

23 St. Machar Drive
Aberdeen
Scotland
AB2 1RY
United Kingdom

Date of deposit	Accession Number
22 November 1993	NCIMB 40600

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

CLAIMS

1. An in vivo expression system comprising a conjugate comprising a GOI and a Ltp2 gene promoter comprising the sequence shown as SEQ. I.D. 1, or a sequence that has substantial homology with that of SEQ. I.D. 1, or a variant thereof; wherein the conjugate is integrated, preferably stably integrated, within a monocotyledon's genomic DNA.

2. A transgenic cereal comprising a conjugate comprising a GOI and a Ltp2 gene promoter as defined in claim 1 wherein the conjugate is integrated, preferably stably integrated, within a cereal's genomic DNA.

3. The in vivo expression in the aleurone cells of a monocotyledon of a conjugate comprising a GOI and a Ltp2 gene promoter as defined in claim 1; wherein the conjugate is integrated, preferably stably integrated, within the monocotyledon's genomic DNA.

4. A method of enhancing in vivo expression of a GOI in just the aleurone cells of a monocotyledon which comprises stably inserting into the genome of those cells a DNA conjugate comprising a Ltp2 gene promoter as defined in claim 1 and a GOI; wherein in the formation of the conjugate the Ltp2 gene promoter is ligated to the GOI in such a manner that each of the myb site and the myc site in the Ltp2 gene promoter is maintained substantially intact.

5. Use of a myb site and a myc site in a Ltp2 gene promoter to enhance in vivo expression of a GOI in just in the aleurone cells of a monocotyledon wherein the Ltp2 gene promoter and the GOI are integrated into the genome of the monocotyledon.

6. A conjugate comprising a GOI and a Ltp2 gene promoter comprising the sequence shown as SEQ. I.D. 1, or a sequence that has substantial homology with that of SEQ. I.D. 1, or a variant thereof.

7. The invention of any one of claims 1 to 6 wherein the promoter is a barley aleurone specific pr moter.

8. The invention of claim 7 wherein the promoter is for a 7 kDa lipid transfer protein.

9. The invention of any one of claims 1 to 8 wherein the promoter is used for expression of a GOI in a cereal seed.

10. The invention of any one of claims 1 to 9 wherein the promoter is used for expression of a GOI in a transgenic cereal seed.

11. The invention of any one of claims 1 to 10 wherein the cereal seed is any one of a rice, maize, wheat, or barley seed, preferably maize.

12. The invention of any one of claims 1 to 11 wherein the promoter is the promoter for Ltp2 of *Hordeum vulgare*.

13. The invention according to any one of the preceding claims wherein the conjugate further comprises at least one additional sequence to increase expression of a GOI or the GOI.

14. The invention according to any one of the preceding claims wherein the conjugate is stably integrated within the genome of a developing grain.

FIGURE 1

-807	GATCTCGATGTGTAGTCTACGAGAAGG
-780	GTTAACCGTCTCTTCGTGAGAATAACCGTGGCCTAAAAATAAGCCGATGAGGATAAATAA
-720	AATGTGGTGGTACAGTACTTCAAGAGGTTTACTCATCAAGAGGATGCTTTTCCGATGAGC
-660	TCTAGTAGTACATCGGACCTCACATACCTCCATTGTGGTGAAATATTTTGTGCTCATTTA
-600	GTGATGGGTAAATTTTGTTTATGTCACTCTAGGTTTTGACATTTCAGTTTTGCCACTCTT
-540	AGGTTTTGACAAATAATTTCCATTCCGCGGCAAAAGCAAAACAATTTTATTTTACTTTTA
-480	CCACTCTTAGCTTTCACAATGTATCACAAATGCCACTCTAGAAATTCTGTTTATGCCACA
-420	GAATGTGAAAAAAAACACTCACTTATTTGAAGCCAAGGTGTTCATGGCATGGAAATGTGA
-360	CATAAAGTAACGTTCGTGTATAAGAAAAAATTGTACTCCTCGTAACAAGAGACGGAAACA
-300	TCATGAGACAATCGCGTTTGGAAGGCTTTGCATCACCTTTGGATGATGCGCATGAATGGA
-240	GTCGTCTGCTTGCTAGCCTTCGCCTACCGCCCACTGAGTCCGGGCGG***CAACTA***CCATCGG
-180	CGAACGACC***CAGCTG***ACCTCTACCGACCGGACTTGAATGCGCTACCTTCGTCAGCGACGA
-120	TGGCCGCGTACGCTGGCGACGTGCCCCCG***CATGCATG***GCGGCACATGGCGAGCTCAGACC
- 60	GTGCGTGGCTGGCTACAAATACGTACCCCGTGAGTGCCCTAGCTAGAAACTTACACCTGC
1	AACTGCGAGAGCGAGCGTGTGAGTGTAGCCGAGTAGATCACCGTACGACGACGACGAGG
60	GGC**ATGGCGATGGCGATGGGGATGGCGATGAGGAAGGAGGCAGCGGTGGCCGTGATGATG**
120	**GTGATGGTGGTGACGCTGGCGGCGGGTGCGGACGCGGGAGCGGGAGCGGCGTGCGAGCCG**
180	**GCGCAGCTGGCGGTGTGCGCGTCGGCGATCCTGGGCGGGACGAAGCCGAGCGGCGAGTGC**
240	**TGCGGGAACCTGCGGGCGCAGCAGGGGTGCTTGTGCCAGTACGTCAAGGACCCCAACTAC**
300	**GGGCACTACGTGAGCAGCCCACACGCGCGCGACACCCTCAACTTGTGCGGCATACCCGTA**
360	**CCGCACTGCTAG**CCGCCTAGCCGATCGAGGGCTCCAGGCACGCATGCATGTTCCTGTTAT
420	GTGTATGTTGGAATAAAATGCTGGTGATCTATGGCGGCTAGCTTGCTTCCTGGCTAGCAG
480	CTGCTGTAATGAAATTTGTGTTGCAACTTTTTTTTTTAGTCC

FIGURE 2A

-807 GATCTCGATGTGTAGTCTACGAGAAGG

-780 GTTAACCGTCTCTTCGTGAGAATAACCGTGGCCTAAAAATAAGCCGATGAGGATAAATAA

-720 AATGTGGTGGTACAGTACTTCAAGAGGTTTACTCATCAAGAGGATGCTTTTCCGATGAGC

-660 TCTAGTAGTACATCGGACCTCACATACCTCCATTGTGGTGAAATATTTTGTGCTCATTTA

-600 GTGATGGGTAAATTTTGTTTATGTCACTCTAGGTTTTGACATTTCAGTTTTGCCACTCTT

-540 AGGTTTTGACAAATAATTTCCATTCCGCGGCAAAAGCAAAACAATTTTATTTTACTTTTA

-480 CCACTCTTAGCTTTCACAATGTATCACAAATGCCACTCTAGAAATTCTGTTTATGCCACA

-420 GAATGTGAAAAAAAACACTCACTTATTTGAAGCCAAGGTGTTCATGGCATGGAAATGTGA

-360 CATAAAGTAACGTTCGTGTATAAGAAAAAATTGTACTCCTCGTAACAAGAGACGGAAACA

-300 TCATGAGACAATCGCGTTTGGAAGGCTTTGCATCACCTTTGGATGATGCGCATGAATGGA

-240 GTCGTCTGCTTGCTAGCCTTCGCCTACCGCCCACTGAGTCCGGGCGG***CAACTA***CCATCGG

-180 CGAACGACC***CAGCTG***ACCTCTACCGACCGGACTTGAATGCGCTACCTTCGTCAGCGACGA

-120 TGGCCGCGTACGCTGGCGACGTGCCCCCG***CATGCATG***GCGGCACATGGCGAGCTCAGACC

-060 GTGCGTGGCTGGC$\boxed{\text{TACAAA}}$TACGTACCCCGTGAGTGCCCTAGCTAGAAACTTACACCTGC

FIGURE 2B

-807 GATCTCGATGTGTAGTCTACGAGAAGG

-780 GTTAACCGTCTCTTCGTGAGAATAACCGTGGCCTAAAAATAAGCCGATGAGGATAAATAA

-720 AATGTGGTGGTACAGTACTTCAAGAGGTTTACTCATCAAGAGGATGCTTTTCCGATGAGC

-660 TCTAGTAGTACATCGGACCTCACATACCTCCATTGTGGTGAAATATTTTGTGCTCATTTA

-600 GTGATGGGTAAATTTTGTTTATGTCACTCTAGGTTTTGACATTTCAGTTTTGCCACTCTT

-540 AGGTTTTGACAAATAATTTCCATTCCGCGGCAAAAGCAAAACAATTTTATTTTACTTTTA

-480 CCACTCTTAGCTTTCACAATGTATCACAAATGCCACTCTAGAAATTCTGTTTATGCCACA

-420 GAATGTGAAAAAAAACACTCACTTATTTGAAGCCAAGGTGTTCATGGCATGGAAATGTGA

-360 CATAAAGTAACGTTCGTGTATAAGAAAAAATTGTACTCCTCGTAACAAGAGACGGAAACA

-300 TCATGAGACAATCGCGTTTGGAAGGCTTTGCATCACCTTTGGATGATGCGCATGAATGGA

-240 GTCGTCTGCTTGCTAGCCTTCGCCTACCGCCCACTGAGTCCGGGCGG***CAACTA***CCATCGG

-180 CGAACGACC***CAGCTG***ACCTCTACCGACCGGACTTGAATGCGCTACCTTCGTCAGCGACGA

-120 TGGCCGCGTACGCTGGCGACGTGCCCCCG***CATGCATG***GCGGCACATGGCGAGCTCAGACC

- 60 GTGCGTGGCTGGC$\boxed{\text{TACAAA}}$TACGTACCCCGTGAGTGCCCTAGCTAGAAACTTACACCTGC

1 AACTGCGAGAGCGAGCGTGTGAGTGTAGCCGAGTAGATC

FIGURE 3

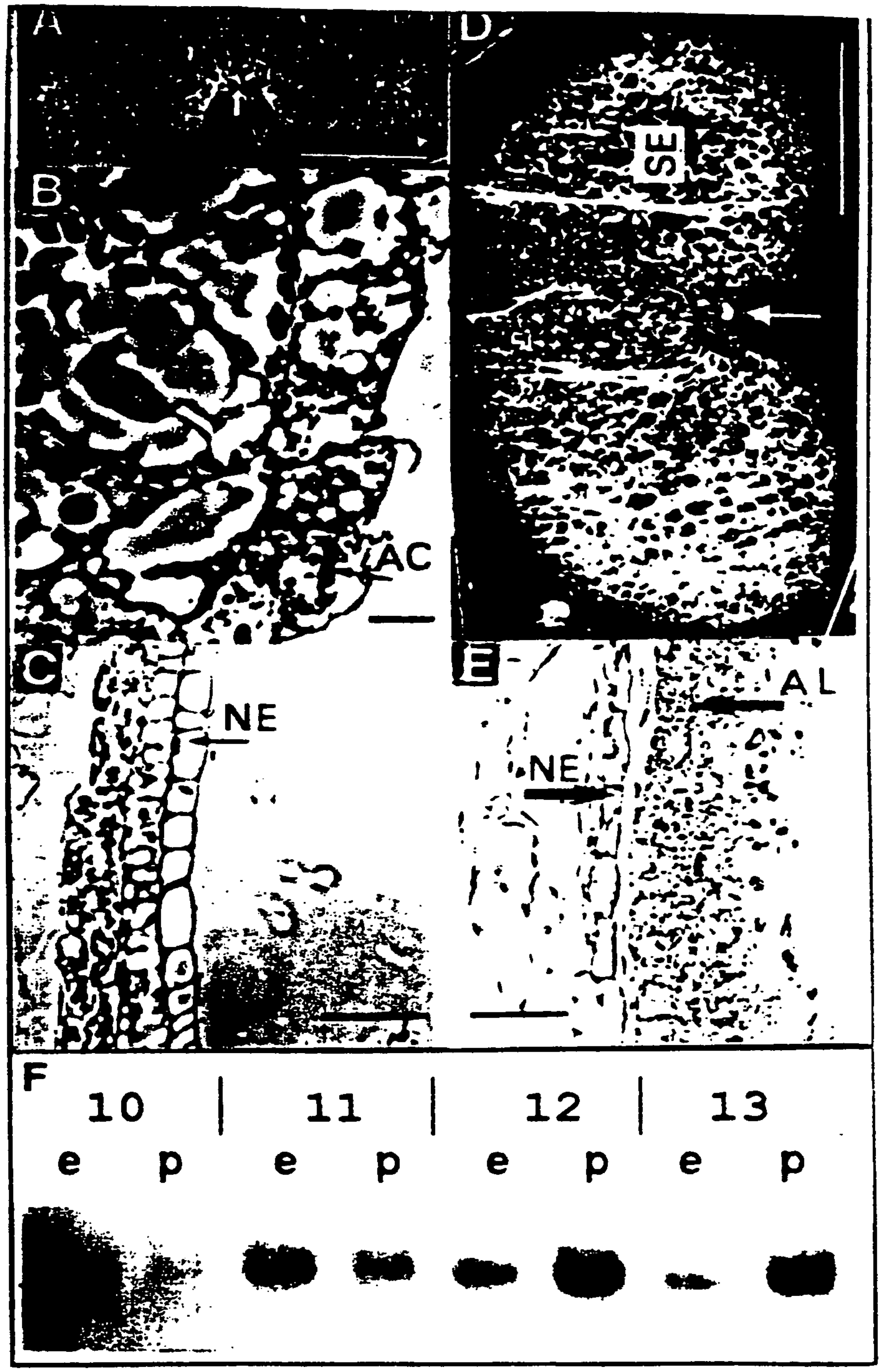


FIGURE 4a-b

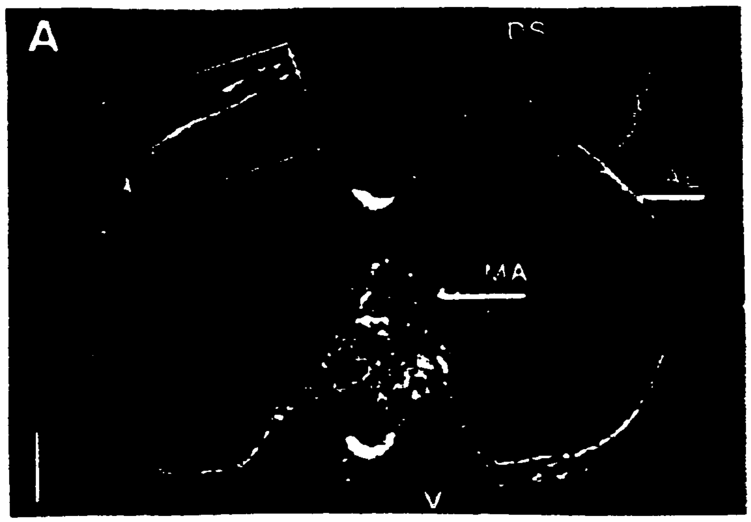


FIGURE 4c-e

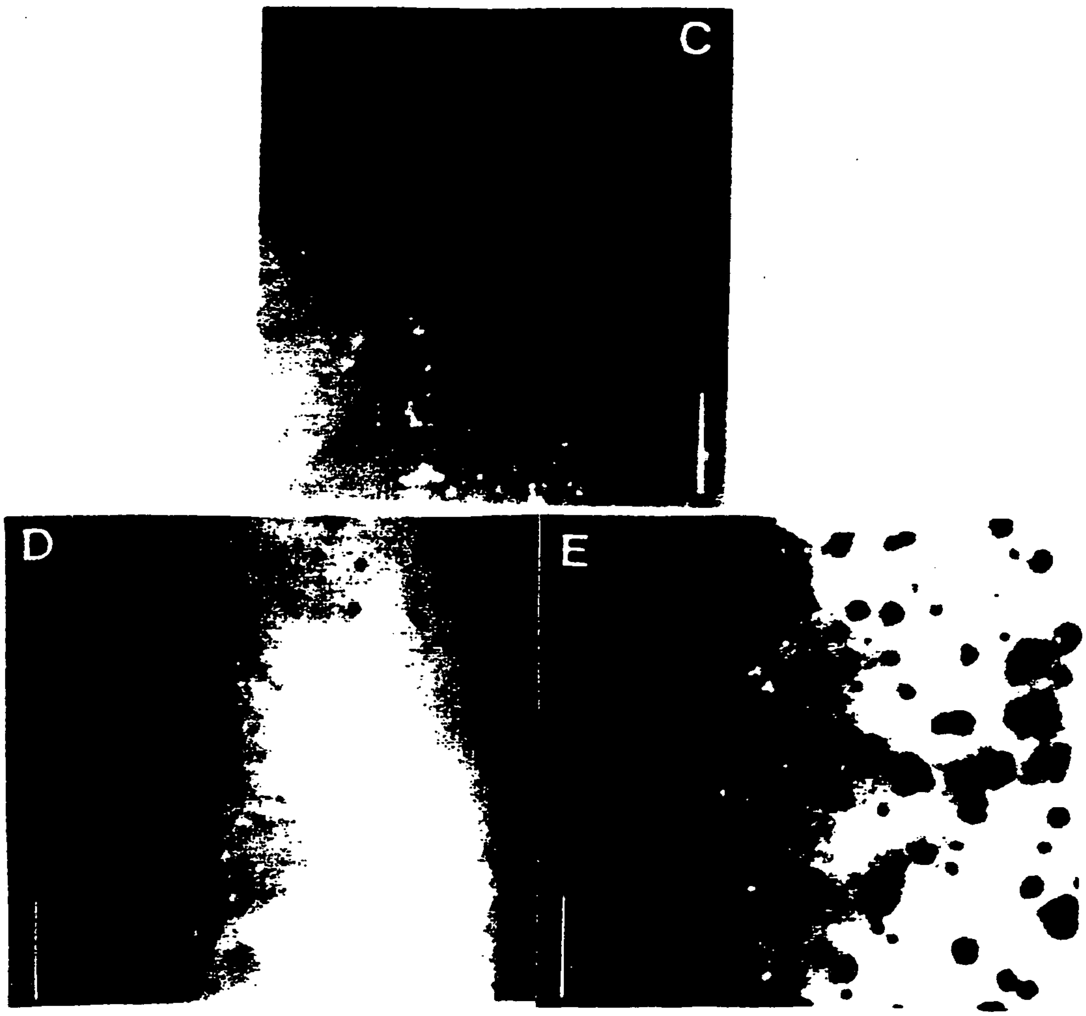


FIGURE 5

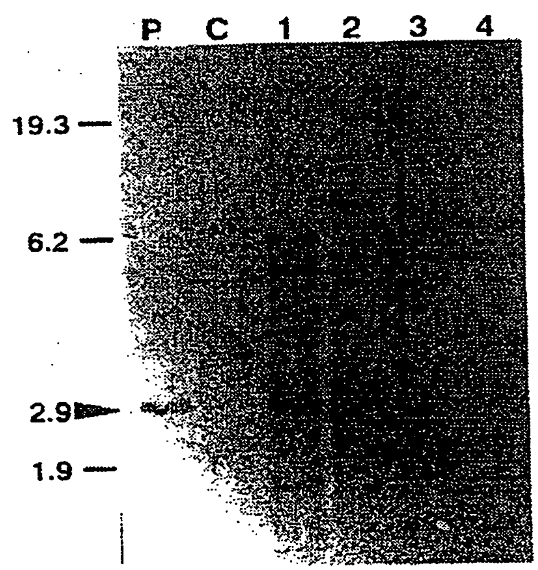


FIGURE 6

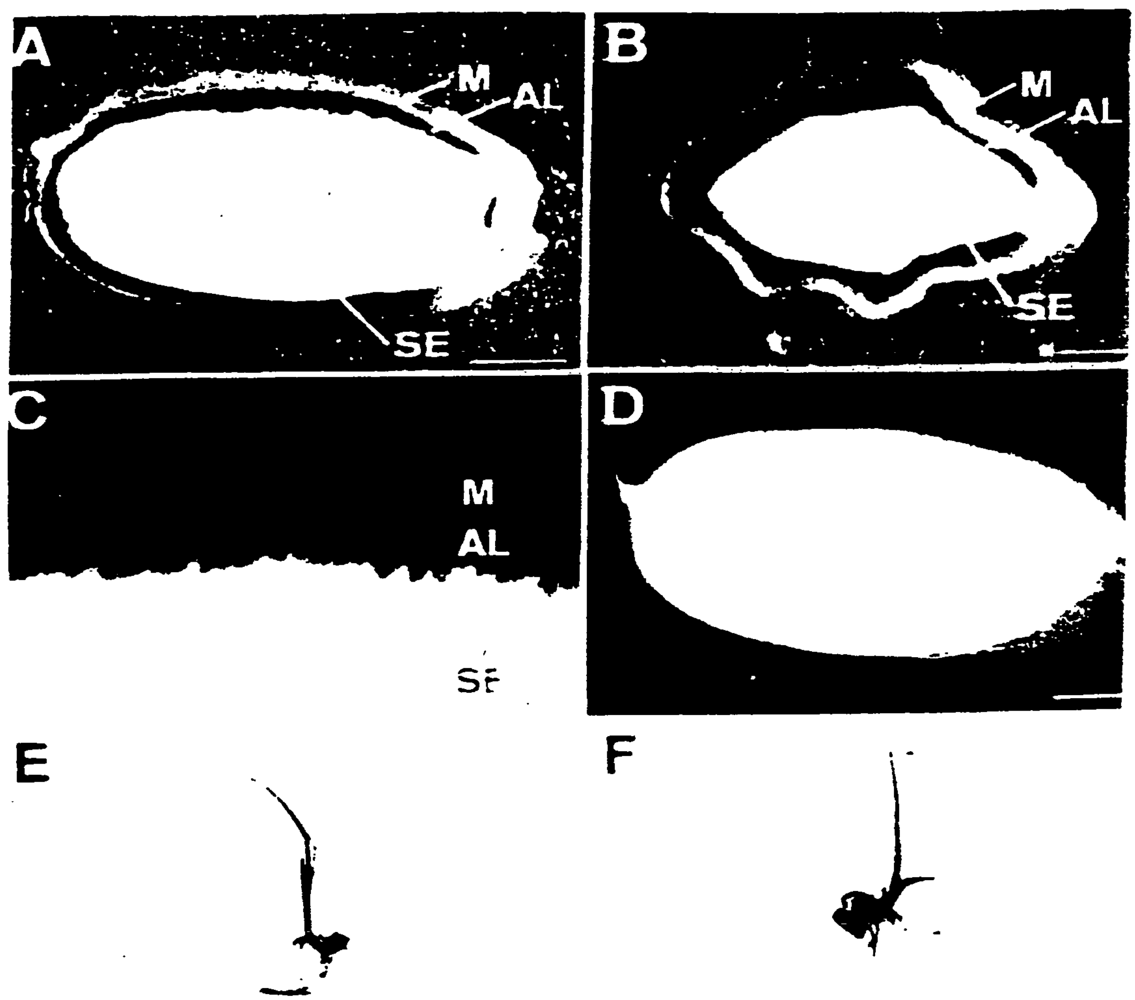


FIGURE 7

```
     MYB                    MYC
    TAACTG                 CANNTG
    C   G
Ltp2 GGCAACTACCATCGGCGAACGACCCAGCTGACCTCTACCGACCGGACTTG- 98nt-TACAAA
```

60

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/82, A01H 5/00 | A1

(11) International Publication Number: WO 95/23230

(43) International Publication Date: 31 August 1995 (31.08.95)

(21) International Application Number: PCT/NO95/00042

(22) International Filing Date: 23 February 1995 (23.02.95)

(30) Priority Data:
9403512.8 24 February 1994 (24.02.94) GB

(71)(72) Applicants and Inventors: OLSEN, Odd-Arne [NO/NO]; Tamveien 16, N-1430 Ås (NO). KALLA, Roger [AU/AU]; Mowle Place, Weetangera, ACT 2641 (AU). LINNESTAD, Casper [NO/NO]; Ovre Linnestad, N-1540 Vestby (NO).

(74) Agent: A/S BERGEN PATENTKONTOR; Strandgt. 191, N-5000 Bergen (NO).

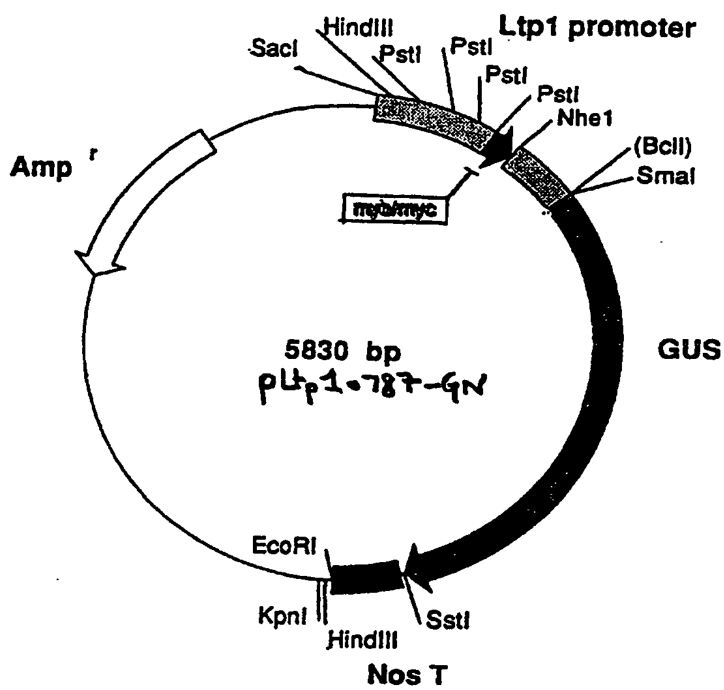
(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PROMOTER FROM A LIPID TRANSFER PROTEIN GENE

(57) Abstract

An expression system for at least the aleurone cells of a developing caryopsis or for at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or developing grain or plant (e.g. in the root, leaves and stem) is described. The expression system comprises a gene promoter fused to a GOI (gene of interest). In a preferred embodiment the expression system comprises the GOI fused to a modified Ltp1 gene promoter.



PROMOTER FROM A LIPID TRANSFER PROTEIN GENE

The present invention relates to a promoter and to a construct comprising the same.

In particular the present invention relates to the use of a promoter for the expression of a gene of interest (GOI) in a specific tissue or tissues of a plant.

More in particular the present invention relates to a modified promoter for a lipid transfer protein (Ltp) gene known as the Ltp1 gene. The present invention also relates to the application of this modified Ltp1 gene promoter to express a GOI in a specific tissue or specific tissues of a plant. For example, expression can be in either the aleurone layer or the scutellar epithelial layer of a monocotyledon, especially a transgenic cereal caryopsis (or grain), more especially a developing transgenic cereal caryopsis (or grain). Particular examples include expression in the scutellar epithelial tissue or vascular tissue of a transgenic rice plant, in particular in the vascular bundles and tip of emerging shoots and roots, leaf veins and vascular bundles of stems.

A diagrammatic illustration of a developing caryopsis (or grain) is presented in Figure 1, which is discussed in detail later. In short, a typical developing caryopsis (or grain) comprises an endosperm component and an embryo component. The endosperm, which is the site of deposition of different storage products such as starch and proteins, supports the growth of the emerging seedling during a short period of time after germination. The embryo gives rise to the vegetative plant. These components and aspects are further discussed in Bosnes *et al.* 1992 and Olsen *et al.* 1992.

The embryo component can be divided into a scutellum and an embryo axis. The scutellum can be sub-divided into an epithelial layer, which is usually one cell thick, and an inner body of parenchyma cells. Likewise, the embryo axis can be sub-divided into a root component and a shoot component.

The endosperm component of mature grains can be divided into a peripheral layer of living aleurone cells surrounding a central mass of non-living starchy endosperm cells. The aleurone layer in barley is three cells thick. During caryopsis germination, the cells of the aleurone layer produce amyolytic and proteolytic enzymes that degrade the storage compounds into metabolites that are taken up and are used by the growing embryo.

Two aspects of aleurone cell biology that have been intensively studied are the genetics of anthocyanin pigmentation of aleurone cells in maize (McClintock, 1987) and the hormonal regulation of gene transcription in the aleurone layer of germinating barley caryopsis (Fincher, 1989). Using transposon tagging, several structural and regulatory genes in the anthocyanin synthesis pathway have been isolated and characterized (Paz-Ares *et al.*, 1987; Dellaporta *et al.*, 1988). In barley, α-amylase and β-glucanase genes that are expressed both in the aleurone layer and embryos of mature germinating caryopsis have been identified (Karrer *et al.*, 1991; Slakeski and Fincher, 1992). In addition, two other cDNAs representing transcripts that are differentially expressed in the aleurone layers of developing barley grains have been isolated. These are CHI26 (Lea *et al.*,1991) and pZE40 (Smith *et al.*, 1992).

None of these references discloses expression of those gene products in specific cell types of developing grains of transgenic cereal plants or in the scutellar epithelial tissue or vascular tissue of a germinating rice seedling or a developing rice grain or rice plant.

In the life of a developing caryopsis (or grain), the embryo component of a dried caryopsis will imbibe water. The presence of water triggers the production of the hormone gibberellic acid in the embryo. In barley and other grass caryopsis, the embryo releases the gibberillic acid which in turn causes expression of a number of genes in the aleurone layer of the endosperm resulting in the production of a number of enzymes such as α-amylases, proteases and β-glucanases. Similar enzymes are also produced by expression of genes in the epithelial layer.

These degradative enzymes digest certain components of the developing caryopsis (or grain) to form sugars and amino acids.

For example, the α-amylases digest the starch store in the starchy endosperm, whereas the proteases digest the storage proteins and the β-glucanases digest the cell walls. The resultant sugars and amino acids cross the epithelial layer and trigger growth of the shoot and root of the embryo axis - i.e. start the germination process.

In some cases it is desirable to transform seeds, grains, caryopsis and plants by introducing genes which, as a result of their expression, yield new or improved properties to the resulting transformed seeds, grains, caryopsis or plants. For example, it may be desirable to alter the expression levels of a natural structural gene which may be under- or over- expressed. It may even be desirable to reduce or eliminate a disease which harms or destroys the seed, grain, caryopsis or the plant.

It may even be desirable to make the seed, grain, caryopsis or the plant resistant to herbicides. It may even be desirable to prevent or to reduce the extent of pre-harvest sprouting.

It may even be desirable for the seed, grain, caryopsis, or plant to produce compounds useful for mammalian usage, such as human insulin.

Some techniques are known for addressing some of those aims.

For example, the bacterium *Agrobacterium tumefaciens* has been used to introduce desired genes into the chromosome of a plant. For example the gene coding for EPSP synthase, a key enzyme in the synthesis of aromatic acids in plants, has been isolated and introduced into petunia plants under the control of a CaMV promoter (Shah *et al.*, [1986]). The transgenic plants expressed increased levels of EPSP synthase in their chloroplasts and were more tolerant to glycophophate - which inhibits production of EPSP synthase.

Other examples may be found in R.W. Old & S.B. Primrose (1993). Another use of *Agrobacterium tumefaciens* is described in De Silva *et al.* (1992) wherein a recombinant DNA construct is described containing a plant plastid specific promoter that expresses a gene placed under its control in concert with the fatty acid or lipid biosynthesis in the plant cell.

PCT WO 90/01551 mentions the use of the aleurone cells of mature, germinating caryopsis to produce proteins from GOIs under the control of an α-amylase promoter. This promoter is active only in germinating caryopsis.

Non-specific lipid transfer proteins (nsLtps) have the ability to mediate *in vitro* transfer of radiolabelled phospholipids from liposomal donor membranes to mitochondrial acceptor membranes (Kader *et al.*, 1984; Watanabe and Yamada, 1986). Although their *in vivo* function remains unclear, nsLTPs from plants have recently received much attention due to their recurrent isolation as cDNA clones representing developmentally regulated transcripts expressed in several different tissues. A common feature is that, at some point in their development, they are highly expressed in tissues producing an extracellular layer rich in lipids.

In particular, transcripts corresponding to cDNAs encoding 10 kDa nsLTPs have been characterized in the tapetum cells of anthers as well as the epidermal layers of leaves and shoots in tobacco (Koltunow *et al.*, 1990; Fleming *et al.*, 1992), and barley aleurone layers (Mundy and Rogers, 1986; Jakobsen *et al.*, 1989).

In addition, a 10 kDa nsLTP has been discovered to be one of the proteins secreted from auxin-treated somatic carrot embryos into the tissue culture medium (Sterk *et al.*, 1991).

Based on *in situ* hybridisation data demonstrating that the Ltp transcripts are localized in the protoderm cells of the somatic and zygotic carrot embryo, it was suggested that *in vivo* nsLTPs are involved in either cutin biosynthesis or in the biogenesis and degradation of storage lipids (Sossountzov *et al.*, 1991; Sterk *et al.*, 1991).

A nsLTP in *Arabidopsis* has been localized to the cell walls lending further support to an extracellular function of this class of proteins (Thoma *et al.*, 1993).

Recently, using a standard *in vitro* Ltp assay, two 10 kDa nsLtps and one member of a novel class of 7 kDa nsLtp's were isolated from wheat seeds (Monnet, 1990; Dieryck *et al.*, 1992).

The sequence of this 7 kDa wheat nsLtp protein shows a high degree of similarity with the predicted protein from the open reading frame (ORF) of the Bz11E cDNA, which had been isolated in a differential screening for barley aleurone specific transcripts (Jakobsen *et al.*, 1989). However, the amino acid sequence of this polypeptide showed only limited sequence identities with the previously sequenced 10 kDa proteins. In sub-cellular localisation studies using gold labelled antibodies one 10 kDa protein from *Arabidopsis* was localised to the cell wall of epidermal leaf cells. The presence of a signal peptide domain in the N-terminus of the open reading frames of all characterised plant nsLtp cDNAs, also suggests that these are proteins destined for the secretory pathway with a possible extracellular function.

Olsen *et al.* in a paper titled "Molecular Strategies For Improving Pre-Harvest Sprouting Resistance In Cereals" published in 1990 in the published extracts from the Fifth International Symposium On Pre-Harvest Sprouting In Cereals (Westview Press Inc.) describe three different strategies for expressing different "effector" genes in the aleurone layer in developing grains of transgenic plants. This document mentions 4 promoter systems - including a system called B11E.

Kalla *et al.* (1993) in a paper titled "Characterisation of Promoter Elements Of Aleurone Specific Genes From Barley" describe the possibility of the expression of anti-sense genes by the use of promoters of the aleurone genes B22E, B23D, B14D, and B11E.

Linnestad *et al.* (1991) describe the isolation and sequencing of the Ltp1 gene and disclose a 787 base pair fragment of the Ltp1 gene promoter fused to a fragment of the Ltp1 structural gene. This paper does not disclose any expression studies using the 787 base pair fragment.

Skriver *et al.* (1992) report further on the Ltp1 gene. This paper says that the Ltp1 gene promoter is only aleurone specific. To confirm this submission the paper further reports on the isolation and fusion of a 769 bp fragment (-702 to +67 bp) of the gene to the bacterial β-glucuronidase (*GUS*) reporter gene. This fragment therefore contains 635 bp of the Ltp1 gene promoter. Subsequent transient expression studies showed that the shortened gene promoter resulted only in aleurone specific expression. Expression was not observed in any other tissue. The authors conclude that there are sequences between the -702 and +67 bp of Ltp1 which contain DNA elements that specifically modulate its transcription in aleurone cells.

One of the major limitations to the molecular breeding of new types of crop plants with specific cells expressing GOIs is the lack of a suitable tissue specific promoter. In particular, there is a lack of a tissue specific promoter that leads to expression of a GOI in a developing caryopsis (or grain) or in a germinating rice seedling or in a developing grain, in particular in the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant.

Moreover, all of the available promoters - such as the CaMV 35S, rice actin and maize alcohol dehydrogenase - are constitutive, i.e. they are fairly non-specific in target site or stage development as they drive expression in most cell types in the plants.

Hence, another problem that arises is how to achieve expression of a product coded for by a GOI in a specific tissue that gives minimal interference with the developing embryo and seedling.

Our co-pending United Kingdom patent application (GB 9324707.0) describes the use of an Ltp2 gene promoter for expression of a GOI in the aleurone layer. However, in spite of this teaching, there is still a need for other tissue specific promoters, such as another aleurone specific promoter or, preferably, a promoter specific for vascular tissue and/or the scutellar epithelial layer. In this regard, it is still desirable to provide other tissue specific expression of GOIs in cereals such as rice, maize, wheat, barley and other transgenic cereal plants. Moreover it is desirable to provide tissue specific expression that does not detrimentally affect the developing embryo and the developing caryopsis (or grain).

According to a first aspect of the present invention there is provided a modified Ltp1 gene promoter which is integrated, preferably stably integrated, within a plant material's genomic DNA and which is capable of inducing expression of a GOI when fused to the gene promoter in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem).

According to a second aspect of the present invetion there is provided a modified Ltp1 gene promoter according to claim 1 or claim 2 wherein the promoter comprises the nucleic acid sequence shown as SEQ. I.D. 1, or a sequence that has substantial homology with that of SEQ. I.D. 1, or a variant thereof.

According to a third aspect of the present invetion there is provided an isolated Ltp1 gene promoter comprising the sequence shown as SEQ. I.D. 1, or a sequence that has substantial homology therewith, or a variant thereof.

According to a fourth aspect of the present invetion there is provided a construct comprising a GOI and a modified Ltp1 gene promoter according to the present invention; wherein the construct is capable of being expressed in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of a plant material; and wherein if there is expression in just the aleurone layer of a developing barley caryopsis then the fused promoter and GOI are not the 769 bp fragment of Skriver

et al (1992).

According to a fifth aspect of the present invetion there is provided an expression system for at least the aleurone cells or for at least the scutellar epithelial tissue or vascular tissue of a plant material, the expression system comprising a GOI fused to a modified Ltp1 gene promoter wherein the expression system is capable of being expressed in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of the plant material; and wherein if there is expression in just the aleurone layer of a developing barley caryopsis then the fused promoter and GOI are not the 769 bp fragment of Skriver *et al* (1992).

According to a sixth aspect of the present invetion there is provided an expression system for at least the aleurone cells of a developing caryopsis or for at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or developing grain or plant (e.g. in the root, leaves and stem), the expression system comprising a gene promoter fused to a GOI wherein the expression system is capable of being expressed in at least the aleurone cells of the developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem); either wherein if there is expression in just the aleurone layer of a developing barley caryopsis then either the promoter is not the wild type Ltp1 promoter in its natural enviroment and the GOI is not the Ltp1 functional gene in its natural enviroment; or wherein if there is expression in just the aleurone layer of a developing caryopsis then the fused promoter and GOI are not the 769 bp fragment of Skriver *et al* (1992).

According to a seventh aspect of the present invetion there is provided a transgenic cereal comprising an expression system according to the present invention or a construct according to the present invention wherein the expression system or construct is integrated, preferably stably integrated, within the cereal's genomic DNA.

According to an eighth aspect of the present invetion there is provided the use of a gene promoter according to the present invention to induce expression of a GOI when fused to the gene promoter in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of a plant material.

According to a ninth aspect of the present invetion there is provided a process of expressing a GOI when fused to a gene promoter according to the present invention wherein expression occurs in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of a plant material.

According to a tenth aspect of the present invetion there is provided a process of expressing in at least the scutellar epithelial tissue or vascular tissue of a developing grain or a germinating seedling or a plant, preferably a developing rice grain or a germinating rice seedling or a transgenic rice plant, an expression system according to teh present invention or a construct according to the present invention wherein the expression system or construct is integrated, preferably stably integrated, within the cereal's genomic DNA.

According to an eleventh aspect of the present invetion there is provided a combination expression system comprising a. as a first construct, a construct according to the present invention; and b. as a second construct, a construct comprising a GOI and another gene promoter that is tissue- or stage-specific.

According to a twelfth aspect of the present invetion there is provided a developing cereal grain, preferably a germinating rice seedling, comprising any one of: a promoter according to the present invention, an expression system according to the present invention, a construct according to the present invention, or a combination expression system according to the present invention.

According to a thirteenth aspect of the present invetion there is provided plasmid NCIMB 40609.

Preferably the plant material is a developing caryopsis, a germinating seedling, a developing grain or a plant.

Preferably the construct is capable of being expressed in at least the aleurone cells of a developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant when the construct is integrated, preferably stably integrated, within the caryopsis's or grain's or seedling's or plant's genomic DNA.

Preferably the modified Ltp1 gene promoter comprises the nucleic acid sequence shown as SEQ. I.D. 1, or a sequence that has substantial homology with that of SEQ. I.D. 1, or a variant thereof.

Preferably the construct further comprises at least one additional sequence to increase expression of the GOI.

Preferably the expression system is for at least the aleurone cells of a developing caryopsis or for at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or developing grain or plant (e.g. in the root, leaves and stem).

Preferably the expression system is additionally capable of being expressed in the embryo cells of the germinating grain or the plantlet.

Preferably the expression system is integrated, preferably stably integrated, within a developing caryopsis's genomic DNA or a germinating seedling's genomic DNA or a developing grain's genomic DNA or a plant's genomic DNA.

Preferably, in the expression system, the gene promoter comprises the sequence shown as SEQ I.D. No. 1 or comprises a sequence that has substantial homology therewith, or is a variant thereof.

Preferably, the expression system comprises the construct according to the present invention.

Preferably, in the use, the gene promoter is used to induce expression of a GOI when fused to the gene promoter in at least the aleurone cells of a developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem).

Preferably, the gene promoter expresses the GOI when fused to the gene promoter in at least the aleurone cells of a developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem).

Preferably the promoter and GOI are integrated, preferably stably integrated, within a cereal's genomic DNA.

Preferably the gene promoter is a fragment of a barley Ltp1 gene promoter.

Preferably the promoter is for a 10 kDa lipid transfer protein.

Preferably the gene promoter is obtainable from plasmid NCIMB 40609.

Preferably the gene promoter is used for expression of a GOI in a cereal caryopsis or a cereal grain or a cereal seedling or a cereal plant.

Preferably the cereal caryopsis is a developing cereal caryopsis, the cereal grain is a developing cereal grain, and the cereal seedling is a germinating cereal seedling.

Preferably the cereal is any one of a rice, maize, wheat, or barley.

Preferably the cereal is rice or maize.

Preferably the developing caryopsis is a developing barley caryopsis, the germinating seedling is a germinating rice seedling, the developing grain is a developing rice grain, and the plant is a transgenic rice plant.

Preferably in the combination expression system each construct is integrated, preferably stably integrated, within a plant material.

Preferably each of the myb site and the myc site in the gene promoter is maintained substantially intact.

Preferably the gene promoter is integrated, preferably stably integrated, in the developing caryopsis's genomic DNA or the germinating seedling's genomic DNA or the developing grain's genomic DNA or the plant's genomic DNA and which is capable of inducing expression of a GOI when fused to the gene promoter in at least the aleurone cells of the developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem).

Preferably the transgenic developing caryopsis, germinating seedling, developing grain or plant is prepared by stable integration of the GOI and the gene promoter to form a stable transgenic plant. This ensures aleurone or epithelial or vascular expression at, at least, the developing caryopsis stage. One preferred method for achieving this includes preparing the transgenic developing caryospis, germinating seedling, developing grain or plant by stable integration of the GOI and the gene promoter at the protoplast level.

Preferably the promoter is used for expression of a GOI in a monocotyledonous species, including a grass - preferably a transgenic cereal grain or caryopsis. Preferably the gene promoter is used for expression of a GOI in a cereal grain or caryopsis. Preferably the cereal grain or caryopsis is a developing cereal grain or caryopsis. Preferably the cereal grain or caryopsis is any one of a rice, maize, wheat, or barley grain or caryopsis.

Preferably the cereal grain is a rice grain.

Preferably the DNA sequence for the modified Ltp1 gene promoter is the nucleic acid sequence shown as SEQ. I.D. 1.

Preferably in the combination expression system each construct is integrated, preferably stably integrated, within a developing caryopsis's genomic DNA or a grain's genomic DNA or a seedling's genomic DNA or a plant's genomic DNA.

Preferably, in the combination expression system, the first construct comprises the modified Ltp1 gene promoter according to the present invention.

Preferably, the promoter in the second construct is an aleurone specific promoter.

Preferably the promoter in the second construct a barley promoter.

Preferably the second construct is the B22E gene promoter.

Preferably the promoter in the second construct is the Ltp2 gene promoter.

Preferably the promoter in the second construct is for a 7 kDa lipid transfer protein.

Preferably the promoter in the second construct is the promoter for Ltp2 of *Hordeum vulgare*.

Preferably the promoter in the second construct comprises the sequence shown as SEQ. I.D. 2, or a sequence that has substantial homology therewith, or a variant thereof.

Preferably each of the myb site and the myc site in the Ltp2 gene promoter is maintained substantially intact.

Preferably the second construct further comprises at least one additional sequence to increase expression of the GOI.

Preferably, in the combination expression system, the grain or caryopsis is as defined above for the present invention.

Preferably the gene promoter is obtainable from plasmid NCIMB 40609.

A preferred embodiment of the present invention is a modified Ltp1 gene promoter which is integrated, preferably stably integrated, within a plant material's genomic DNA and which is capable of inducing expression of a GOI when fused to the gene promoter in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem), but wherein if there is expression in just the aleurone layer of a developing seed then the fused promoter and GOI are not the 769 bp fragment of Skriver *et al* (1992).

An even more preferred embodiment of the present invention is a modified Ltp1 gene promoter which is integrated, preferably stably integrated, within a plant material's genomic DNA and which is capable of inducing expression of a GOI when fused to the gene promoter in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem), wherein the promoter comprises the nucleic acid sequence shown as SEQ. I.D. 1, or a sequence that has substantial homology with that of SEQ. I.D. 1, or a variant thereof.

As a highly preferred embodiment, the present invention therefore provides transgenic rice comprising a construct comprising a GOI fused to a modified Ltp1 gene promoter; wherein the construct is integrated, preferably stably integrated, within the rice's genomic DNA, and wherein the GOI is expressed in at least the vascular tissue and/or scutellar epithelial layer of a germinating rice seedling or a developing rice grain or a rice plant.

In a more preferred embodiment the present invention provides a transgenic rice seedling, grain or plant comprising a construct comprising a GOI fused to a modified Ltp1 gene promoter, wherein the construct is integrated, preferably stably integrated, within the rice's genomic DNA; wherein the GOI is expressed in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant, and wherein the modified Ltp1 gene promoter comprises the nucleic acid sequence shown as SEQ. I.D. 1, or a sequence that has substantial homology with that of SEQ. I.D. 1, or a variant thereof.

The additional sequence(s) for the construct(s) for increasing the expression of the GOI(s) may be one or more repeats (e.g. tandem repeats) of the promoter upstream box(es) which are responsible for the aleurone layer or scutellar epithelial cell and/or vascular expression pattern of the modified Ltp1 gene promoter. The additional sequence may even be a *Sh1*-intron.

The term "plant material" includes a developing caryopsis, a germinating caryopsis or grain, or a seedling, a plantlet or a plant, or tissues or cells thereof, such as the aleurone cells of a developing caryopsis or the scutellar epithelial tissue or vascular tissue of a germinating seedling or developing grain or plant (e.g. in the root, leaves and stem).

Thus a preferred aspect of the present invention comprises plant material comprising a GOI and a modified Ltp1 gene promoter which is capable of inducing expression of the GOI when fused to the gene promoter in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of the plant material; wherein the construct is capable of being expressed in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of the plant material, when the construct is integrated, preferably stably integrated, within the caryopsis's or grain's or seedling's or plant's genomic DNA; and wherein the modified Ltp1 gene promoter comprises the nucleic acid sequence shown as SEQ. I.D. 1, or a sequence that has substantial homology with that of SEQ. I.D. 1, or a variant thereof.

The term "modified" with reference to the present invention means any Ltp1 gene promoter that is different to the wild type promoter but wherein the promoter induces expression in at least the aleurone cells of a developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem).

In particular, a preferred modified Ltp1 gene promoter is a shortened wild type Ltp1 gene promoter but wherein the promoter induces expression in at least the aleurone cells of a developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem).

The term "transgenic" in relation to the present invention - in particular in relation to the developing caryopsis, germinating seedlings, developing grains and plants of the present invention - does not include a wild type promoter in its natural enviroment in combination with its associated functional gene (GOI) in its natural enviroment. Thus, the term includes developing caryopsis or seedlings or grains or plants incorporating the GOI which may be natural or non-natural to the grain or caryopsis or seedling or grain or plant in question operatively linked to the modified Ltp1 promoter of the present invention.

The term "GOI" with reference to the present invention means any gene of interest. A GOI can be any gene that is either foreign or natural to the cereal in question, except for the wild type Ltp1 functional gene when in its natural enviroment. In the combination expression system the GOI may be the same or different.

Typical examples of a GOI include genes encoding for proteins and enzymes that modify metabolic and catabolic processes. For example, the GOI may be a protein giving added nutritional value to the grain or caryopsis as a food or crop. Typical examples include plant proteins that can inhibit the formation of anti-nutritive factors and plant proteins that have a more desirable amino acid composition (e.g. a higher lysine content than the non-transgenic plant).

The GOI may even code for an enzyme that can be used in food processing such as chymosin, thaumatin, α-galactosidase and guar.

In a preferred embodiment, particularly with vascular expression, the GOI may code for an agent for introducing or increasing pathogen resistance.

The GOI may even be an antisense construct for modifying the expression of natural transcripts present in the relevant tissues.

The GOI may even code for a non-natural plant compound that is of benefit to animals or humans. For example, the GOI could code for a pharmaceutically active protein or enzyme such as the therapeutic compounds insulin, interferon, human serum albumin, human growth factor and blood clotting factors. In this regard, the transformed cereal grain or caryopsis could prepare acceptable quantities of the desired compound which could be easily retrievable from the scutellar epithelial layer, the aleurone layer or the vascular tissue.

Preferably the GOI is a gene encoding for any one of a protein having a high nutritional value, a *Bacillus thuringensis* insect toxin, an α- or β- amylase antisense transcript, a protease antisense transcript, or a glucanase antisense transcript.

The term "a variant thereof" with reference to the present invention means any substitution of, variation of, modification of, replacement of, deletion of or the addition of one or more nucleic acid(s) from or to the promoter sequence providing the resultant sequence exhibits at least aleurone, scutellar epithelial or vascular expression, respectively. The term also includes sequences that can substantially hybridise to the promoter sequence.

The term "substantial homology" covers homology with respect to at least the essential nucleic acids/nucleic acid residues of the promoter sequence providing the homologous sequence acts as a promoter, e.g. as a promoter for at least aleurone expression in a developing caryopsis or for at least scutellar epithelial tissue or

vascular tissue expression in a germinating seedling or in a developing grain or plant. Preferably there is at least about 80% homology, more preferably at least about 90% homology, and even more preferably there is at least about 95% homology with the promoter sequence shown as SEQ. I.D. No. 1. or SEQ. I.D. No. 2, respectively.

The term "maintained substantially intact" means that at least the essential components of each of the myb site and the myc site remain in the construct to ensure acceptable expression of a GOI. Preferably at least about 75%, more preferably at least about 90%, and even more preferably there is at least about 95%, of the myb or myc site is left intact.

The term "construct" - which is synonymous with terms such as "cassette", "hybrid" and "conjugate" - includes a GOI directly or indirectly attached to the modified gene promoter, such as to form a [modified Ltp1 gene promoter-GOI] cassette. An example of an indirect attachment is the provision of a suitable spacer group such as an intron sequence, such as the *Sh1*-intron, intermediate the promoter and the GOI. The same is true for the term "fused" in relation to the present invention which includes direct or indirect attachment.

The term "expression system" means that the system defined above can be expressed in an appropiate organism, tissue, cell or medium. In this regard, the expression system of the present invention may comprise additional components that ensure ro increase the expression of the GOI by use of the gene promoter.

As indicated above, the expression system of the present invention can also be used in conjunction with another expression system, preferably an expression system that is also tissue and/or stage specific.

For example, the construct comprising the modified Ltp1 gene promoter (e.g. the 787 bp fragment of SEQ. I.D. NO. 1) can be used in conjunction with a construct comprising the Ltp2 gene promoter (e.g. SEQ. I.D. NO. 2) - which is the subject of our co-pending UK patent application GB 9324707.0.

In this respect, and with reference to barley, in the early stages of developing caryopsis the modified Ltp1 gene promoter affects expression of a GOI in at least the aleurone layers of developing caryopsis. This expression can then be complimented by use of the Ltp2 gene which can express a GOI (which may be the same or different as that operatively linked to the modified Ltp1 gene promoter) in high levels in the aleurone layer of developing grains.

However, the combination expression system is very effective for transgenic rice. In this respect, in the early stages of developing caryopsis the modified Ltp1 gene promoter expresses a GOI in the scutellar epithelial layer and the vascular tissue. This expression can then be complimented by use of the Ltp2 gene which can express a GOI in high levels in the aleurone layer of developing grains. This combination is particularly advantageous for pre-harvest sprouting when the first response is production of α-amylase in the scutellar epithelium cells as this can be reduced or prevented by placing an anti-sense α amylase gene under the control of the Ltp1 promoter. In this system, the expression of antisense α-amylase would block the synthesis of α-amylases in the scutellum epithelial cells - where they are first made. The same or another GOI could be expressed in the aleurone layer *via* the Ltp2 gene promoter.

The construct comprising the modified Ltp1 gene promoter may even be used in conjunction with a construct comprising the B22E gene promoter - details of which may be found in Olsen *et al.* (1990) and Klemsdal *et al.*, (1991). This gene promoter, which is expressed in immature aleurone layers, has been shown by particle bombardment experiments to be capable of driving *Gus* expression in developing barley grains. Also, using Northern analysis, as well as *in situ* hybridization, it has been shown that the B22E cDNA probe hybridizes to transcripts in the aleurone layer and in the scutellum parenchyma cells and the provascular bundle of the embryo axis in developing barley grains. In addition, a hybridizing transcript is also present in the ventral vascular strand of developing caryopsis (Olsen *et al.*, 1990).

We have also found that by using a 4.6 kb B22E promoter fragment contained on a XbaI-ClaI fragment of a genomic clone fused to the *Gus* reporter gene transformed rice plants could be prepared. Those transformed rice plants exhibited strong expression in the vascular tissue (phloem) of the ventral strand of the developing rice grain. This expression pattern was completely unexpected in view of Klemsdal *et al* (1991). Expression, although weaker, in the same cell type was also observable in the stem of young shoots. Thus, using the B22E promoter, a GOI transcript can be expressed in the aleurone layers of developing grains, the parenchyma cells of the embryonic scutellum and the ventral vascular bundle of developing grains.

The combination of the use of the modified Ltp1 gene promoter and the B22E gene promoter could even include the use of another gene promoter, such as the Ltp2 gene promoter, to express three GOIs respectively wherein each GOI may be the same or different.

One or more of the other expression systems to be used in conjunction with the modified Ltp1 gene promoter expression system may be contained in or on the same transmission vector - such as in the same transforming baterium or even in the same plasmid. The advantage of this is that each expression system can then be delivered at the same time. The respective expression systems will then be turned on during the relevant life time of the grain or caryopsis or the plantlet or the mature plant.

The present invention therefore provides the novel and inventive use of a promoter which can express a GOI in at least the aleurone cells of a developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem). In a preferred embodiment the present invention relates to the use of a modified Ltp1 gene promoter, preferably the Ltp1 gene promoter is obtainable from barley.

The main advantage of the present invention is that the use of the modified Ltp1 gene promoter results in expression of a GOI in at least the aleurone layer of at least a developing caryopsis, such as a developing barley caryopsis, or in at least the

scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or plant of cereals such as rice, maize, wheat or other transgenic cereal grain or caryopsis, preferably a developing rice grain.

Another advantage is that, depending on the type of GOI, the expressed products can be stable *in vivo*. Hence over a period of time high levels of the expressed product can accumulate in the aleurone cells or in epithelial cells or in the vascular tissue.

A further advantage is that the expression of the product coded for by a GOI in the aleurone layer or the epithelial layer or the vascular tissue has minimal interference with the developing embryo and seedling. This is in direct contrast to known constitutive promoters which give high levels of expression in the developing seedling and mature plant tissues which severely affect normal plant development. Thus the present invention is particularly useful for expressing a GOI in at least the aleurone layer of a developing caryopsis or in at least the scutellar epithelial tissue or in the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or plant - such as cereal grains or caryopsis - and in doing so not detrimentaly affect the caryopsis, seedling, grain or plant.

With regard to the first aspect of the present invention it is to be noted that this is the first reported case for the specific expression of a GOI in the scutellar epithelial cells or vascular cells of a transformed developing cereal grain such as rice.

With regard to some aspects of the present invention, it is to be noted that up until now it was believed that the wild type Ltp1 gene promoter or a specific varaint thereof when fused to at least a segment of the Ltp1 fucntional gene would lead only to expression in the aleurone layer. For example see the teachings of Skriver *et al.* (1992). However, with the present invention, we have now surprisingly found that this is not the case and it is now possible to modify the Ltp1 gene promoter to lead to a pronounced expression in at least the aleurone layer or in at least the scutellar epithelial layer or vascular tissue of a plant material.

In one embodiment the plant material is barley plant material. In another embodiment the plant material is not barley plant material. In a preferred embodiment the plant material is rice plant material. In an alternative preferred embodiment the plant material is maize plant material.

In a germinating, transgenic barley caryopsis according to the present invention, there is expression in the aleurone layer.

In a germinating, transgenic rice seedling according to the present invention there is pronounced expression in the scutellar epithelial tissue and vascular tissue.

As indicated, the expression pattern for the present invention is particularly surprising as it was completely unexpected that a modified Ltp1 gene promoter could result in expression of a GOI, such as a plant functional gene, in the aleurone cells of, for example, barley or in the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or plant of rice (see experimental section later). The findings of the present invention are also surprisingly different to the work of Skriver *et al.* (1992) who, as mentioned above, report that the Ltp1 gene promoter and a shortened version thereof when fused to the functional Ltp1 functional gene only result in aleurone specific expression in barley - i.e. expression is not observed in any other tissue in barley or even other cereals.

In order to prepare the transgenic organism according to the present invention, the modified Ltp1 gene promoter may be initially inserted into a plasmid. For example, the *Sac*I-*Bcl*I Ltp1 gene promoter fragment can be inserted into the *Sac*I-*BamH*I site of Bluescript. A GOI, such as *GUS*, can then be inserted into this construct. Furthermore, a *Sh*1 intron can then be inserted into the *Sma*I site of this construct.

Stable integration into protoplasts may be achieved by using the method of Shimamoto (1989). Another way is by bombardment of an embryonic suspension of cells (e.g. rice, barley or maize cells). A further way is by bombardment of immature embryos (e.g. rice, maize or barley embryos).

With regard to the present invention, it is shown by using particle bombardments that the modified Ltp1 gene promoter, such as the 787 bp fragment of the attached sequence, when fused to a β-glucuronidase (*GUS*) reporter gene, which serves as a GOI for the purposes of this invention, acts as a promoter for expression of *GUS* in a specific tissue type or specific tissue types. For example, *GUS* expression can be achieved in the aleurone cells of developing cereal caryopsis or grain, in particular developing barley caryopsis, or in the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or plant, in particular developing rice grain or germinating seedlings.

In particular, in transgenic rice plants, the modified barley Ltp1 gene promoter directs strong expression of the *GUS*-reporter gene in the scutellar epithelial layer and the vascular tissue of the developing caryopsis. This expression can continue through into the germinating grains. The surprising finding is that very pronounced expression can be seen in the scutellar epithelial tissue or vascular tissue of a developing rice grain or germinating rice seedlings. Other examples include expression in the vascular bundles and tip of emerging shoots and roots, leaf veins and vascular bundles of stems.

Generally therefore the present invention relates to a modified promoter for a Ltp1 gene encoding a 10 kDa nsLTP. In the present invention, a genomic clone was isolated using the cDNA insert of previously isolated cDNA clone and characterised by DNA sequencing (see discussion later). The sequence of the cDNA and isolated genomic clone was found to be identical in the overlapping region. It was found the Ltp1 gene contains one intron (see discussion later).

By comparing the DNA sequence of the active promoter sequences two putative *cis*-acting elements with the potential of binding known transcriptional factors present in cereals were detected. They include the binding sites for transcriptional factors of the myb and myc class, namely TAACTG and CANNTG respectively. Our studies showed that high levels of expression are achieved when the myb and myc sites are left intact.

In the present invention, mature fertile rice plants were regenerated from transformed cultured rice protoplasts. The developing caryopsis of these primary transformants were analysed for the expression of *GUS*. It was found that the modified barley Ltp1 gene promoter confers some expression in the aleurone layer of the transgenic rice plants. However, pronounced expression was observed in the scutellar epithelial tissue or vascular tissue of germinating rice seedlings or developing transgenic rice grain or transgenic rice plants. This is the first example of such patterns of expression in transgenic rice plants.

The following sample has been deposited in accordance with the Budapest Treaty at the recognised depositary The National Collections of Industrial and Marine Bacteria Limited (NCIMB) at 23 St Machar Drive, Aberdeen, Scotland, AB2 1RY, United Kingdom, on 11 January 1994:

> An *E. Coli* K12 bacterial stock containing the plasmid pLtp1.787-GN - i.e. Bluescript containing a 787 bp fragment of the barley Ltp1 gene promoter (Deposit Number NCIMB 40609).

The plasmid pLtp1.787-GN is shown pictorially in Figure 6 (see later).

The modified Ltp1 gene promoter can be isolated from this plasmid through the use of appropriate PCR primers, which may be easily constructed from the data from the shown sequences.

Other embodiments and aspects of the present invention include:

> A transformed host having the capability of expressing a GOI in the aleurone layer or the scutellar epithelial layer or the vascular tissue through the use of the gene promoter as hereinbefore described;

> A vector incorporating a construct as hereinbefore described or any part thereof;

A plasmid comprising a construct as hereinbefore described or any part thereof;

A cellular organism or cell line transformed with such a vector;

A monocotylenedonous plant comprising any one of the same;

A developing caryopsis or grain or germinating seedling comprising any of the same; and

A method of expressing any one of the above.

The present invention will now be described only by way of examples in which reference shall be made to the accompanying Figures in which:

Figure 1 is a diagrammatic illustration of the structural components of a developing caryopsis;

Figure 2 shows the results for an *in situ* hybridization experiment for a wild type Ltp1 gene promoter in barley;

Figure 3 is a nucleotide sequence of part of the wild type Ltp1 gene taken from Linnestad *et al.* (1991);

Figure 4 is a nucleotide sequence of part of the wild type Ltp1 gene taken from Skriver *et al.* (1992);

Figure 5 is a nucleotide sequence of a 787 bp fragment of the wild type Ltp1 gene promoter;

Figure 6 is a linear map of the Ltp1.787-GN construct showing additional sequence information;

Figure 7 is a circular map of the plasmid pLtp1.787-GN containing the Ltp1.787-GN construct;

Figure 8 is a longitudinal section of a developing rice grain post expression of the modified Ltp1 gene promoter; and

Figure 9 is a longitudinal section of a mature germinating rice grain post expression of the modified Ltp1 gene promoter.

A. METHODS

i. Plant material

Caryopsis of barley (*Hordeum vulgare* cv. Bomi) were collected from plants grown in a phytotron as described before (Kvaale and Olsen, 1986). The plants were emasculated and pollinated by hand and isolated in order to ensure accurate determination of caryopsis age.

ii. cDNA and genomic clones

The isolation and sequencing of the Ltp1 cDNA clone was conducted as described by Jakobsen *et al.* (1989). A barley, cv. Bomi genomic library was constructed by partial *Mbo*I digestion of total genomic DNA and subsequent ligation of the 10-20 kilo basepair (kb) size fraction with *Bam*HI digested lambda EMBL3 DNA (Clontech Labs, Palo Alto, Ca, USA). Using the Ltp1 cDNA insert as a template for probe synthesis with a random labelling kit (Boehringer-Mannheim), one positive clone was identified after repeated rounds of plaque hybridization. DNA purified from this clone was restricted with several enzymes and characterized by Southern blot analysis. The sequence data obtained after this procedure are shown in Figure 3.

iii. *In situ* hybridization

For *in vitro* transcription of antisense RNA, the plasmid Ltp1 was linearized and transcribed using MAXIscript (Ambion) and [α ^{33}P]-UTP (Amersham International). The probe was hydrolysed to fragments of about 100 bp as described by Somssich *et al.* (1988). Caryopsis tissues were fixed in 1% glutaraldehyde, 100 mM sodium phosphate (pH 7.0) for 2 hours and embedded in Histowax (Histolab, Göteborg, Sweden).

Barley caryopsis sections of 10 μm were pre-treated with pronase (Calbiochem) as described by (Schmelzer *et al.*, 1988) and hybridized with 25 ml of hybridization mix (200 ng probe ml-1, 50% formamide, 10% (w/v) dextran sulphate, 0.3 M NaCl, 10 mM Tris-HCl, 1 mM EDTA (pH 7), 0.02% polyvinyl-pyrrolidone, 0.2% Ficoll, 0.02% bovine serum albumin) for 15 hours at 50 °C.

Post-hybridization was carried out according to Somssich *et al.* (1988) and autoradiography was done as described by Schmelzer *et al.* (1988).

iv. Constructs for transient expression analysis

For the micro-projectile bombardment experiments, the following was used:

pLtp1.787-GN (see Figure 7 and associated commentary).

Isolated plasmid DNA was used in the bombardment studies.

For transient assay studies with rice protoplasts, the following were studied:

pLtp1.787-GN (see Figure 7 and associated commentary).

pLtp1.787(*-myb/myc*)-GN (see commentary below).

Deletion studies were performed on the modified Ltp1 gene promoter (Ltp1.787) wherein a section of DNA containing the myb and myc sites (see Figure 3 and associated commentary) was removed to form pLtp1.787(*-myb/myc*)-GN. In this embodiment, the modified Ltp1 gene promoter having deletions from and between the myb and myc sites was prepared and fused to GN. In order to prepare this deleted modified Ltp1 gene promoter a PCR strategy using primers covering the flanking sequences of the deleted sequence was adopted.

v. Transformation of barley cells by particle bombardment

Barley caryopsis were harvested at 25 DAP (days after pollination), surface sterilized in 1% sodium hypochlorite for 5 min and then washed 4 times in sterile distilled water. The maternal tissues were removed to expose the aleurone layer and the caryopsis was then divided into two, longitudinally along the crease. The pieces of tissue were then placed, endosperm down, onto MS media (Murashige & Skoog 1962) with 10 g/l sucrose solidified with 10 g/l agar in plastic petri dishes (in two rows of 4 endosperm halves per dish). Embryos from the same caryopsis were placed in the same petri dishes with the scutellum side facing upwards.

Single bombardments were performed in a DuPont PDS 1000 device, with M-17 tungsten pellets (approx. 1 μm in diameter) coated with DNA as described by Gordon-Kamm *et al.* (1990) and using a 100 mm mesh 2 cm below the stopping plate. Histochemical staining for *GUS* expression was performed with X-Gluc (5-bromo,4-chloro,3-indolyl,ß-D,Glucuronic acid) as described by Jefferson (1987) at 37°C for 2 days.

In these studies, after bombardment with the pLTp1.787-GN and staining for *GUS*-activity, blue spots appeared both in the aleurone layer as well as in the scutellar epithelium layer. These results demonstrate that the 787 bp fragment of the Ltp1 gene promoter of the present invention is capable of driving transcription in the epithelial cells.

vi. Rice transformation

In these studies, the gene was transformed into rice by electroporation of embryogenetic protoplasts following the teachings of Shimamoto et al. 1989. Six fertile transgenic rice plants were obtained. Histochemical *GUS* analysis was also carried out with developing rice grains of 25 DAP and 1 to 5 day old seedings and up to 1 month old plants derived from transgenic grains. The results demonstrated expression of the Ltp1 - *GUS* gene in the scutellar epithelial layer of developing transgenic rice plants. In addition, in a germinating rice seedling according to the present invention there is a pronounced expression in the vascular tissue.

B. RESULTS AND DISCUSSION WITH REFERENCE TO THE FIGURES

1. In order to explain more fully the results, reference is made to Figure 1 which shows the major components of a typical developing caryopsis (or grain) 1. In this regard, the caryopsis (or grain) 1 comprises an endosperm component 3 and an embryo component 5. The endosperm component 3 is divisible into an outer aleurone layer 7, which is three cells thick for barley caryopsis, and a starchy endosperm 9. The embryo component 5 is divisible into a scutellum 11 and an embryo axis 13. The scutellum 11 is further divisible into an epithelial layer 15 and parenchyma layer 17. Likewise, the embryo axis 13 is further divisible into a root component 19 and a shoot component 21.

2. Figure 2 is a transverse section of a 30 day-old wild-type developing barley caryopsis showing *in situ* hybridisation with a radio-labelled Ltp1 probe. The bound probe is only seen in the aleurone layer. It is not seen in any other tissue type, in particular the scutellar epithelial layer. This work confirms the work of Skriver *et al.* (1992).

The bright spots are due to optical interference.

3. Figure 3 shows the nucleotide sequence and the deduced amino acid sequence of Ltp1. The intron is indicated by lower case letters. The TGA stop codon is indicated by an asterisk, the putative CAAT and TATA sequences are indicated by boxes. A 21 bp inverted repeat is indicated by arrows. Four 8 bp palindromic sequences are overlined. The motif indicated by thick underlining resembles the CATGTAAA motif present in the promoters of several genes expressed in aleurone cells (Klemsdal *et al.* (1991)). An AT block followed by a myb consensus recognition site and a myc binding motif are indicated by double underlining.

4. Figure 4 shows the sequence of the Ltp1 gene. The 351 bp open reading frame is interrupted by a 133 bp intron (+412 to +544). The transcript start site is at position +1. The putative CAAT and TATA boxes are at -107 and -34. A putative poly (dA) site is at +785 (Skriver *et al.* (1992)).

5. Figure 5 is the nucleotide sequence of the preferred embodiment of the present invention, i.e. a 787 bp fragment of the Ltp1 gene promoter. The same commentary for Figure 3 is equally applicable here.

6. Figure 6 is an outline of the Ltp1 genomic clone containing the Ltp1 structural gene (shaded box) and the promoter fragment fused to the *GUS* gene (black box) used to transform rice. Also indicated are the extensions of the Ltp1 fragment described in Linnestad *et al.* (1991) and Skriver *et al.* (1992). The figures used represent DNA fragment lengths in kb. The total length of the genomic clone is in the order of 8.1 kb.

7. Figure 7 helps explain how pLtp1.787-GN was constructed. In this regard, the following fragments were sequentially cloned into the vector Bluescript KS$^-$: firstly the 787 bp SacI/Bcl fragment of the Ltp1 gene promoter was cloned into the SacI/BclI site of the vector; and secondly a *GUS*-Nos Terminator on 2150 bp SmaI/EcoRI fragment derived from pBI101 was cloned into SmaI/EcoRI downstream of the Ltp1 promoter.

8. Figure 8 is a longitudinal section of a 30 day old transgenic rice grain showing transcriptional activity of the construct of Figure 7 (i.e. pLtp1.787-GN) containing the promoter of Figure 5. It is to be noted that transcriptional activity is achieved in the scutellar epithelial layer, as denoted by the blue staining.

9. Figure 9 is a longitudinal section of a mature germinating transgenic rice grain showing transcriptional activity of the construct of Figure 7 (i.e. pLtp1.787-GN) containing the promoter of Figure 5.

It is to be noted that transcriptional activity is achieved in the scutellar epithelial layer. Transcriptional activity is also observed in the shoot epithelial layer and in the aleurone layer. However, the extent of expression in the last two tissue types is not as pronounced as that in the scutellar epithelial layer.

However, more importantly, with the transgenic rice transcriptional activity is observed in the vascular tissue of the germinating seedling and the vascular tissue of the root and stem.

C. SUMMATION

The Examples relate to the isolation of and to the use of a 787 bp fragment of the promoter for the barley Ltp1 gene, which encodes a 10 kDa nsLTP. The gene was isolated by the use of a cDNA from a differential screening experiment in which the positive probe was constructed from aleurone cell poly (A) rich RNA, and the negative probe from the starchy endosperm of immature grains.

A construct comprising the Ltp1 gene promoter fragment and a GOI (in this case *GUS*) was stably inserted into rice protoplasts.

Expression and *in situ* analysis for the wild type gene promoter demonstrated that the Ltp1 transcript is expressed in high levels only in the aleurone cells in developing barley caryopsis. This expression continued in germinating grains and also in

plantlets and mature plants.

However, for transgenic cereals, especially rice, even though there is some expression in the aleurone layer for the modified Ltp1 gene promoter it is, however, not as pronounced as that in each of the epithelial cells of the scutellum, the epidermal cells of the coleoptile and the vascular strands of the embryo of developing caryopsis (or grain).

This result is completely unexpected as it shows that a modified Ltp1 promoter can function differently in transgenic cereals, especially rice, than the wild-type Ltp1 gene in barley.

Expression and histochemical analysis for the transgenic rice demonstrated that the Ltp1 transcript is expressed in high levels in the scutellar epithelial tissue and vascular tissue, especially of a germinating rice seedling and a developing rice grain and a rice plant (e.g. in the root, leaves and stem). This expression continued in germinating grains and also in plantlets and mature plants.

Importantly, for rice, expression is observed in the vascular tissue of the germinating seedling and the vascular tissue of the root and stem.

This result is completely unexpected in view of the expression pattern of wild-type Ltp1 gene in barley.

Using the 787 bp promoter fragment in particle bombardments of developing barley caryopsis, we obtained activity (blue spots) in the epithelium layer of the scutellum.

The results therefore indicate that the modified Ltp1 gene promoter directs expression of a GOI predominantly in the aleurone cells of developing caryopsis, particularly for barley, or the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem) particularly for rice.

The modified Ltp1 gene promoter therefore represents a valuable tool for the expression of GOIs in the aleurone layer of developing caryopsis, in particular developing barley caryopsis.

Moreover, the modified Ltp1 gene promoter represents a valuable tool for the expression of GOIs in the scutellar epithelial cells and vascular cells of germinating seedlings or developing grain, in particular developing or germinating rice seedlings or grain. The epithelial or vascular expression is of particular benefit because the 787 bp LTP1 gene fragment can be used to express antisense α-amylase in the scutellar epithelial layer in order to reduce or to prevent damage due to preharvest sprouting or to introduce or enhance pathogen resistance.

One possible reason for the expression activity of the modified Ltp1 gene promoter of the present invention may be the absence of "silencer" elements in the modified gene promoter which prevent expression of the wild type gene in, for example, the scutellar epithelial layer and vascular cells. Accordingly, the term "modified" (as defined above) could include removal of any silencer elements from the wild type Ltp1 gene promoter.

Studies with the modified Ltp1 gene promoter having deletions from and between the myb and myc sites when fused to GN showed that the relative activity of the deleted modified Ltp1 gene promoter was less (in some cases 70% less) than the modified Ltp1 gene promoter which contains the myb and myc sites. Therefore, it is believed that the presence of the myb and myc sites are important for even higher levels of expression of the modified Ltp1 promoter in at least protoplasts of at least rice.

Accordingly the present invention also covers a method of enhancing the *in vivo* expression of a GOI in at least the aleurone layer of a developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant preferably of an embryo of a developing monocotyledon grain or caryopsis, comprising stably inserting into the genome of those cells a DNA construct comprising a modified Ltp1 gene promoter and a GOI, wherein in the

formation of the construct the modified Ltp1 gene promoter is ligated to the GOI in such a manner that each of the myb site and the myc site in the modified Ltp1 gene promoter is maintained substantially intact.

The present invention also covers the use of a myb site and a myc site in a modified Ltp1 gene promoter to enhance *in vivo* expression of a GOI in at least the aleurone layer of a developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant, preferably of an embryo of a developing monocotyledon caryopsis or grain, wherein the modified Ltp1 gene promoter and the GOI are integrated into the genome of the monocotyledon.

Each of these aspects is applicable to the combination expression system.

D. CONCLUSIONS VIS-A-VIS THE SPECIFIC EXAMPLES

1. The barley Ltp1 gene encodes a protein homologous to the 10 kDa wheat lipid transfer protein.

2. The wild type Ltp1 gene promoter is expressed in developing barley aleurone cells.

3. The modified Ltp1 gene promoter is transiently expressed in developing barley scutellar epithelial cells after particle bombardment.

4. The modified Ltp1 gene promoter directs expression of the *GUS*-reporter gene in the scutellar epithelial cells of developing transgenic rice grains. However, more pronounced expression is observed in the vascular tissue of germinating seedlings and the root and stem of the transgenic rice plant.

5. The modified Ltp1 gene promoter contains sequence elements implicated in the transcriptional control of cereal endosperm specific genes.

6. The modified Ltp1 gene promoter contains myb and myc sequence elements that are implicated in the level of transcription in cereal endosperm.

Other modifications of the present invention will be apparent to those skilled in the art without departing from the scope of the invention.

REFERENCES

Bosnes, M., Weideman, F. and Olsen, O.-A. (1992) Endosperm differentiation in barley wild-type and sex mutants. *Plant J.* **2**, 661-674.

De Silva *et al.* 1992 *WO 92/18634*

Dellaporta, S.L., Greenblatt, I., Kermicle, J.L., Hick, J.B. and Wessler, S. (1988) Molecular cloning of the *R-nj* allelel by transposon tagging with *Ac*. In *Chromosome structure and function: Impact of new concepts*, 18th Stadler Genetics Symposium (Gustafson, J.P. and Appels, R., eds.), New York: Plenum Press, pp. 263-282.

Dieryck, W., Gautier, M.-F., Lullien, V. and Joudrier, P. (1992) Nucleotide sequence of a cDNA encoding a lipid transfer protein from wheat (*Triticum durum* Desf.). *Plant Mol. Biol.* **19**, 707-709.

Fincher, G.B. (1989) Molecular and cellular biology associated with endosperm mobilization in germinating cereal grains. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 305-346.

Fleming, A.J, Mandel, T., Hofmann, S., Sterk, P., de Vries, S.C., and Kuhlemeier, C. (1992) Expression pattern of a tobacco lipid transfer protein gene within the shoot apex. *Plant J.* **2**, 855-862.

Gordon-Kamm, W.J., Spencer, T.M., Mangano, M.L., Adams, T.R., Daines, R.J., Start, W.G., O'Brien, J.V., Chambers, S.A., Adams, W.R., Willetts, N.G., Rice, T.B., Mackey, C.J., Krueger, R.W., Kausch, A.P. and Lemaux, P.G. (1990) Transformation of maize cells and regeneration of fertile transgenic plants. *Plant Cell* **2**, 603-618.

Jacobsen, J.V., Knox, R.B. and Pyliotis, N.A. (1971) The structure and composition of aleurone grains in the barley aleurone layer. *Planta* **101**, 189-209.

Jakobsen, K., Klemsdal, S., Aalen, R., Bosnes, M., Alexander, D. and Olsen, O.-A. (1989) Barley aleurone cell development: molecular cloning of aleurone-specific cDNAs from immature grains. *Plant Mol. Biol.* **12**, 285-293.

Jefferson, R.A. (1987) Assaying chimeric genes in plants: the *GUS* gene fusion system. *Plant Mol. Biol. Rep.* **5**, 387-405.

Kader, J.-C., Julienne, M. and Vergnolle, C. (1984) Purification and characterisation of a spinach-leaf protein capable of transferring phospholipids from liposomes to mitochondria or chloroplasts. *Eur. J. Biochem.* **139**, 411-416.

Karrer, E.J., Litts, J.C and Rodriguez, R.L. (1991) Differential expression of α-amylase genes in germinating rice and barley seeds. *Plant Mol. Biol.* **16**, 797-805.

Klemsdal, S.S., Hughes, W., Lonneborg, A., Aalen, R. and Olsen, O.-A. (1991) Primary structure of a novel barley gene differentially expressed in immature aleurone layers. *Mol. Gen. Genet* **228** 9-16.

Koltunow, A.M., Truettner, J., Cox, K.H., Wallroth, M. and Goldberg, R.B. (1990) Different temporal and spatial expression patterns occur during anther development. *Plant Cell* **2**, 1201-1224.

Kosugi, S., Ohashi, Y., Nakajima, K. and Arai, Y. (1990) An improved assay for β-glucuronidase in transformed cells: methanol almost completely suppresses a putative endogenous β-glucuronidase activity. *Plant Sci.* **70**, 133-140.

Kvaale, A. and Olsen, O.-A. (1986) Rates of cell division in developing barley endosperms. *Ann. Bot.* **57**, 829-833.

Lea, R., Tommerup, H., Svendsen, I. and Mundy, J. (1991) Biochemical and molecular characterization of three barley seed proteins with antifungal properties. *J. Biol. Chem.* **266**, 1564-73.

Linnestad, C., Lönneborg, A., Kalla, R. and Olsen, O.-A. (1991) The promoter of a lipid transfer protein gene expressed in barley aleurone cells contains similar Myb and Myc recognition sites as the maize *Bz-McC* allele. *Plant Physiol.* **97**, 841-843.

McClintock, B. (1978) Development of the maize endosperm as revealed by clones. In *The clonal basis of development* (Subtelny, S. and Sussex, I.M., eds.), New York: Academic Press, pp.217-237.

Monnet, F.-P. (1990) *Ph.D thesis.* Universite des Sciences et Techniques du Languedoc, Montpellier, France, pp. 121.

Mundy, J. and Rogers, J. (1986) Selective expression of a probable amylase/protease inhibitor in barley aleurone cells: comparison to the barley amylase/subtilisin inhibitor. *Planta* **169**, 51-62.

Murashige, T., and Skoog, F. (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473-497.

Old & S.B. Primrose (1993) *Principles of Gene Manipulation - An Introduction to Genetic Engineering 4th Edition.* Pub. Blackwell Scientific Publications. Pages 360-363.

Olsen, O.-A., Jakobsen, K.S. and Schmelzer, E. (1990) Development of barley aleurone cells: temporal and spatial patterns of accumulation of cell specific mRNAs. *Planta* **181** 462-466

Olsen, O.-A., Potter, R.H. and Kalla, R. (1992) Histo-differentiation and molecular biology of developing cereal endosperm. *Seed Sci. Res.* **2**, 117-131.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A. and Saedler, H. (1987) The regulatory *c1* locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J.* **6**, 3553-3558.

Schmelzer, E., Jahnen, W. and Hahlbrock, K. (1988) *In situ* localization of light-induced chalcone synthetase mRNA, chalcone synthetase, and flavonoid end products in epidermal cells of parsley leaves. *Proc. Natl. Acad. Sci. U.S.A* **85**, 2989-2993.

Shah *et al.* [1986] *Science* **233** 478-81

Shimamoto, K., Terada, R., Izawa, T. and Fujimoto, H. (1989) Fertile transgenic rice plants regenerated from transformed protoplasts. *Nature* **338**, 274-276.

Skriver, K., Leah, R., Müller-Uri, F., Olsen, F.-L. and Mundy, J. (1992) Structure and expression of the barley lipid transfer gene Ltp1. *Plant Molecular Biology* **18** 585-589.

Slakeski, N. and Fincher, G.B. (1992) Developmental regulation of (1-3,1-4)-b-glucanase gene expression in barley. *Plant Physiol.* **99**, 1226-1231.

Smith, L.M., Handley, J., Li, Y., Martin, H., Donovan, L. and Bowles, D.J. (1992) Temporal and spatial regulation of a novel gene in barley embryos. *Plant Mol. Biol.* **20**, 255- 266.

Somssich, I.E., Schmelzer, E., Kawalleck, P. and Hahlbrock, K. (1988) Gene structure and *in situ* transcript localization of the pathogenesis-related protein 1 in parsley. *Mol. Gen. Genet.* **213**, 93-98.

Sossountzov, L., Riuz-Avila, L., Vignois, F., Jolliot, A., Arondel, V., Tchang, F., Grosbois, M., Guerbette, F., Miginiac, E., Delsney, M., Puigdomenech, P. and Kader, J.-C. (1991) Spatial and temporal expression pattern of a maize lipid transfer protein gene. *Plant Cell* **3**, 923-933.

Sterk, P., Booij, H., Schellekens, G.A., Van Kammen, A. and De Vries, S.C. (1991) Cell-specific expression of the carrot EP2 lipid transfer protein gene. *Plant Cell* **3**, 907-921.

Thoma,S., Kaneko,Y. and Sommerville,C. (1993) A non-specificlipid transfer protein from Arabidopsis is a cell wall protein. *The Plant Journal* **3(3)**, 427-436.

Watanabe, S. and Yamada, M. (1986) Purification and characterization of a non-specific lipid transfer protein from germinated castor bean endosperms which transfers phospholipids and galactolipids. *Biochim. Biophys. Acta.* **876**, 116-123.

SEQUENCE LISTING

(1) GENERAL INFORMATION

NAME OF APPLICANTS: O.-A. OLSEN AND R. KALLA
BUSINESS ADDRESS: PLANT MOLECULAR BIOLOGY LABORATORY
DEPARTMENT OF BIOTECHNICAL SCIENCES
AGRICULTURAL UNIVERSITY OF NORWAY
AND AGRICULTURAL BIOTECHNOLOGY PROGRAM NRC
NORWAY
N-1432

TITLE OF INVENTION: PROMOTER

(2) INFORMATION FOR SEQUENCE I.D. 1

SEQUENCE TYPE:	NUCLEIC ACID
MOLECULE TYPE:	DNA (GENOMIC)
ORIGINAL SOURCE:	BARLEY
SEQUENCE LENGTH:	787
STRANDEDNESS:	DOUBLE
TOPOLOGY:	LINEAR
SEQUENCE:	

-787	GAGCTCC	AAGGCATCAC	CAAGCTTCTA	TGACGCCAAA
-750	ACATCCAAGA	AAGATATGTA	CTAGGATACC	AAGCACCCAA
	GAGTAAACGG	AGGAAGTATA	ATATAAGGCC	CTGTTTGATA
-670	ACAAAGTAGT	AAAAAAACTA	AAGTATTAAA	AACTGCAGTA
	ATTTTACGTG	TAGATAGAAA	ATACCATGGT	TTTAATATAA
-590	TAATATTTTT	TGCAGTATTC	ACAATGTAGA	GAAACTGTTT
	GATTACGCCA	CATATTACTG	CAGTTTAGAT	CGAGCAAGTA
-510	CACGGGAAGA	AGATAACGAC	GTCCCACCCC	TTCTTTTCGC
	CTTCTCTGTT	TTTTAAAAAG	AGGTCTGGGG	TTAGTTTTTT
-430	CAATACTGCA	GTTTTAAAAT	CACAATTCTT	AGAGGCAACC
	AAACACCTCA	TTGTAAATAA	AACTATGATA	ATCTCCAAAA
-350	CTGCAGTATT	CTAAAAATAC	TACAAAAATT	CTTTGTTATC
	AAACAGGGCC	TAAGGAGTTA	AAAAAATTTA	GCCGTAACTG
-270	AGACTCGGCG	AGGCACCAGC	AGCTAGCAGT	CATCAACACT
	TGATGGTTGG	CAAAGCCGAG	TCGACGTGTC	GCGGGGCTCG
-190	GCCTGAGCGG	GAGATACAAT	CTGTTCTCCA	GTAACCCCGT
	CGATTTGGCC	CGCCGACTAA	AGCATCCAGG	CATCTCTCGC
-110	TCGAACCCCT	ATTTAAGCCC	CTCCATTCCT	CCCAACATTC
	TCCACACCTC	CACGAGTTGC	TCATCACTAG	CTAGTACGTT
-30	GTACTGTTAG	CTACAGATTA	AGAAGTGATC	

NOTE: ABOVE SEQUENCE IS A RETYPED VERSION OF FIGURE 5 WHICH IS TO BE TAKEN AS THE CORRECT SEQUENCE

(3) INFORMATION FOR SEQUENCE I.D. 2

SEQUENCE TYPE:	NUCLEIC ACID
MOLECULE TYPE:	DNA (GENOMIC)
ORIGINAL SOURCE:	BARLEY
SEQUENCE LENGTH:	-807
STRANDEDNESS:	DOUBLE
TOPOLOGY:	LINEAR
SEQUENCE:	

-807		GATCTCG	ATGTGTAGTC	TACGAGAAGG
-780	GTTAACCGTC	TCTTCGTGAG	AATAACCGTG	GCCTAAAAAT
	AAGCCGATGA	GGATAAATAA	AATGTGGTGG	TACAGTACTT
	CAAGAGGTTT	ACTCATCAAG	AGGATGCTTT	TCCGATGAGC
-660	TCTAGTAGTA	CATCGGACCT	CACATACCTC	CATTGTGGTG
	AAATATTTTG	TGCTCATTTA	GTGATGGGTA	AATTTTGTTT
	ATGTCACTCT	AGGTTTTGAC	ATTTCAGTTT	TGCCACTCTT
-540	AGGTTTTGAC	AAATAATTTC	CATTCCGCGG	CAAAAGCAAA
	ACAATTTTAT	TTTACTTTTA	CCACTCTTAG	CTTTCACAAT
	GTATCACAAA	TGCCACTCTA	GAAATTCTGT	TTATGCCACA
-420	GAATGTGAAA	AAAAACACTC	ACTTATTTGA	AGCCAAGGTG
	TTCATGGCAT	GGAAATGTGA	CATAAAGTAA	CGTTCGTGTA
	TAAGAAAAAA	TTGTACTCCT	CGTAACAAGA	GACGGAAACA
-300	TCATGAGACA	ATCGCGTTTG	GAAGGCTTTG	CATCACCTTT
	GGATGATGCG	CATGAATGGA	GTCGTCTGCT	TGCTAGCCTT
	CGCCTACCGC	CCACTGAGTC	CGGGCGGCAA	CTACCATCGG
-180	CGAACGACCC	AGCTGACCTC	TACCGACCGG	ACTTGAATGC
	GCTACCTTCG	TCAGCGACGA	TGGCCGCGTA	CGCTGGCGAC
	GTGCCCCCGC	ATGCATGGCG	GCACATGGCG	AGCTCAGACC
-60	GTGCGTGGCT	GGCTACAAAT	ACGTACCCCG	TGAGTGCCCT
	AGCTAGAAAC	TTACACCTGC		

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 24, lines 15-17

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution

The National Collections of Industrial and Marine Bacteria Limited (NCIMB)

Address of depositary institution *(including postal code and country)*

23 St. Machar Drive
Aberdeen
Scotland
AB2 1RY
United Kingdom

Date of deposit	Accession Number
11. JAN. 1994	NCIMB 40609

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet ☐

In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

CLAIMS

1. A modified Ltp1 gene promoter which is integrated, preferably stably integrated, within a plant material's genomic DNA and which is capable of inducing expression of a GOI when fused to the gene promoter in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem).

2. A modified Ltp1 gene promoter according to claim 1 wherein the plant material is a developing caryopsis, a germinating seedling, a developing grain or a plant and wherein the gene promoter is integrated, preferably stably integrated, in the developing caryopsis's genomic DNA or the germinating seedling's genomic DNA or the developing grain's genomic DNA or the plant's genomic DNA and which is capable of inducing expression of a GOI when fused to the gene promoter in at least the aleurone cells of the developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem).

3. A modified Ltp1 gene promoter according to claim 1 or claim 2 wherein the promoter comprises the nucleic acid sequence shown as SEQ. I.D. 1, or a sequence that has substantial homology with that of SEQ. I.D. 1, or a variant thereof.

4. An isolated Ltp1 gene promoter comprising the sequence shown as SEQ. I.D. 1, or a sequence that has substantial homology therewith, or a variant thereof.

5. A construct comprising

a GOI and

a modified Ltp1 gene promoter according to any one of claims 1 to 4;

wherein the construct is capable of being expressed in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of a plant material; and

wherein if there is expression in just the aleurone layer of a developing barley caryopsis then the fused promoter and GOI are not the 769 bp fragment of Skriver *et al* (1992).

6. A construct according to claim 5 wherein the construct is capable of being expressed in at least the aleurone cells of a developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant when the construct is integrated, preferably stably integrated, within the caryopsis's or grain's or seedling's or plant's genomic DNA.

7. A construct according to claim 5 or claim 6 wherein the modified Ltp1 gene promoter comprises the nucleic acid sequence shown as SEQ. I.D. 1, or a sequence that has substantial homology with that of SEQ. I.D. 1, or a variant thereof.

8. The construct according to any one of claims 5 to 7 wherein the construct further comprises at least one additional sequence to increase expression of the GOI.

9. An expression system for at least the aleurone cells or for at least the scutellar epithelial tissue or vascular tissue of a plant material, the expression system comprising

a GOI fused to a modified Ltp1 gene promoter

wherein the expression system is capable of being expressed in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of the plant material; and

wherein if there is expression in just the aleurone layer of a developing barley caryopsis then the fused promoter and GOI are not the 769 bp fragment of Skriver *et al* (1992).

10. An expression system according to claim 9 wherein the expression system is for at least the aleurone cells of a developing caryopsis or for at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or developing grain or plant (e.g. in the root, leaves and stem).

11. An expression system according to claims 9 or claim 10 wherein the expression system is additionally capable of being expressed in the embryo cells of the germinating grain or the plantlet.

12. An expression system according to any one of claims 9 to 11 wherein the expression system is integrated, preferably stably integrated, within a developing caryopsis's genomic DNA or a germinating seedling's genomic DNA or a developing grain's genomic DNA or a plant's genomic DNA.

13. An expression system according to any one of claims 9 to 12 wherein the gene promoter comprises the sequence shown as SEQ I.D. No. 1 or comprises a sequence that has substantial homology therewith, or is a variant thereof.

14. An expression system for at least the aleurone cells of a developing caryopsis or for at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or developing grain or plant (e.g. in the root, leaves and stem), the expression system comprising

a gene promoter fused to a GOI

wherein the expression system is capable of being expressed in at least the aleurone cells of the developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in

the root, leaves and stem); either

wherein if there is expression in just the aleurone layer of a developing barley caryopsis then either the promoter is not the wild type Ltp1 promoter in its natural enviroment and the GOI is not the Ltp1 functional gene in its natural enviroment; or

wherein if there is expression in just the aleurone layer of a developing caryopsis then the fused promoter and GOI are not the 769 bp fragment of Skriver *et al* (1992).

15. An expression system according to any one of claims 9 to 14 comprising a construct according to any one of claims 5 to 8.

16. A transgenic cereal comprising an expression system according to any one of claims 9 to 15 or a construct according to any one of claims 5 to 8 wherein the expression system or construct is integrated, preferably stably integrated, within the cereal's genomic DNA.

17. The use of a gene promoter as defined in any one of the preceding claims to induce expression of a GOI when fused to the gene promoter in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of a plant material.

18. The use according to claim 17 wherein the gene promoter is used to induce expression of a GOI when fused to the gene promoter in at least the aleurone cells of a developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem).

19. A process of expressing a GOI when fused to a gene promoter as defined in any one of the preceding claims, wherein expression occurs in at least the aleurone cells or in at least the scutellar epithelial tissue or vascular tissue of a plant material.

20. A process according to claim 19 wherein the gene promoter expresses the GOI when fused to the gene promoter in at least the aleurone cells of a developing caryopsis or in at least the scutellar epithelial tissue or vascular tissue of a germinating seedling or a developing grain or a plant (e.g. in the root, leaves and stem).

21. A process according to claim 19 or claim 20 wherein the promoter and GOI are integrated, preferably stably integrated, within a cereal's genomic DNA.

22. A process of expressing in at least the scutellar epithelial tissue or vascular tissue of a developing grain or a germinating seedling or a plant, preferably a developing rice grain or a germinating rice seedling or a transgenic rice plant, an expression system according to any one of claims 9 to 15 or a construct according to any one of claims 5 to 8 wherein the expression system or construct is integrated, preferably stably integrated, within the cereal's genomic DNA.

23. The invention of any one of claims 1 to 22 wherein the gene promoter is a fragment of a barley Ltp1 gene promoter.

24. The invention of claim 23 wherein the promoter is for a 10 kDa lipid transfer protein.

25. The invention of claim 23 or claim 24 wherein the gene promoter is obtainable from plasmid NCIMB 40609.

26. The invention of any one of claims 1 to 15 wherein the gene promoter is used for expression of a GOI in a cereal caryopsis or a cereal grain or a cereal seedling or a cereal plant.

27. The invention of claim 26 wherein the cereal caryopsis is a developing cereal caryopsis, the cereal grain is a developing cereal grain, and the cereal seedling is a germinating cereal seedling.

28. The invention of claim 26 or claim 27 wherein the cereal is any one of a rice, maize, wheat, or barley.

29. The invention of claim 28 wherein the cereal is rice or maize, preferably rice.

30. The invention according to any one of claims 1 to 29 wherein the developing caryopsis is a developing barley caryopsis, the germinating seedling is a germinating rice seedling, the developing grain is a developing rice grain, and the plant is a transgenic rice plant.

31. A combination expression system comprising

a. as a first construct, a construct according to any one of claims 5 to 8; and

b. as a second construct, a construct comprising a GOI and another gene promoter that is tissue- or stage-specific.

32. A combination expression system according to claim 31 wherein each construct is integrated, preferably stably integrated, within a plant material.

33. A combination expression system according to claim 32 wherein each construct is integrated, preferably stably integrated, within a developing caryopsis's genomic DNA or a grain's genomic DNA or a seedling's genomic DNA or a plant's genomic DNA.

34. A combination expression system according to any one of claims 31 to 33 wherein the first construct comprises a modified Ltp1 gene promoter comprising the nucleic acid sequence shown as SEQ. I.D. 1, or a sequence that has substantial homology with that of SEQ. I.D. 1, or a variant thereof.

35. A combination expression system according to any one of claims 31 to 34 wherein the promoter in the second construct is an aleurone specific promoter.

36. A combination expression system according to any one of claims 31 to 35 wherein the promoter in the second construct a barley promoter.

37. A combination expression system according to any one of claims 31 to 35 wherein the second construct is the B22E gene promoter.

38. A combination expression system according to any one of claims 31 to 37 wherein the promoter in the second construct is the Ltp2 gene promoter.

39. A combination expression system according to claim 38 wherein the promoter in the second construct is for a 7 kDa lipid transfer protein.

40. A combination expression system according to claim 38 or 39 wherein the promoter in the second construct is the promoter for Ltp2 of *Hordeum vulgare*.

41. A combination expression system according to any one of claims 31 to 40 wherein the promoter in the second construct comprises the sequence shown as SEQ. I.D. 2, or a sequence that has substantial homology therewith,. or a variant thereof.

42. A combination expression system according to any one of claims 38 to 41 wherein each of the myb site and the myc site in the Ltp2 gene promoter is maintained substantially intact.

43. A combination expression system according to any one of claims 31 to 42 wherein the second construct further comprises at least one additional sequence to increase expression of the GOI.

44. A developing cereal grain, preferably a germinating rice seedling, comprising any one of: a promoter according to any one of claims 1 to 4 or any claim dependent thereon, an expression system according to any one of claims 9 to 15 or any claim dependent thereon, a construct according to any one of claims 5 to 8 or any claim dependent thereon, or a combination expression system according to any one of

claims 31 to 43 or any claim dependent thereon.

45. The invention of any one of the preceding claims wherein each of the myb site and the myc site in the gene promoter is maintained substantially intact.

46. Plasmid NCIMB 40609.

47. A promoter, a construct or an expression system or a combination expression system substantially as described herein with reference to any one of Figures 5 to 9.

Fig 1

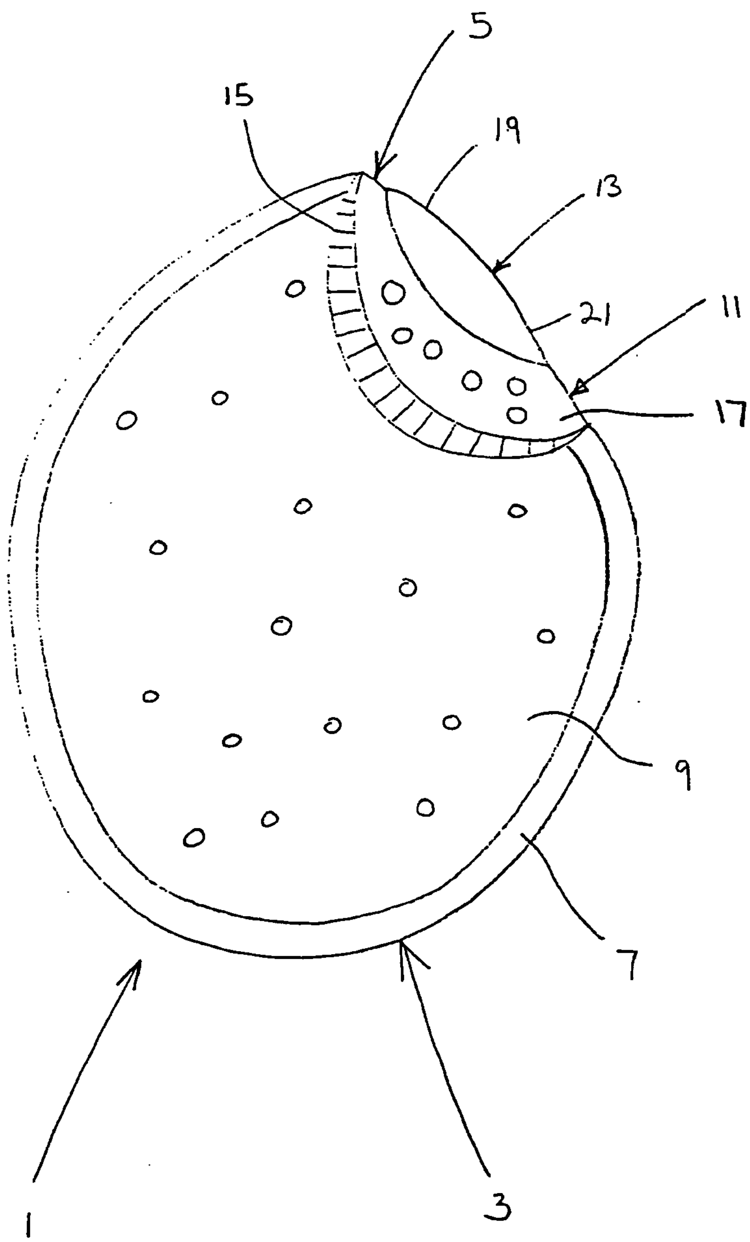


FIG 2

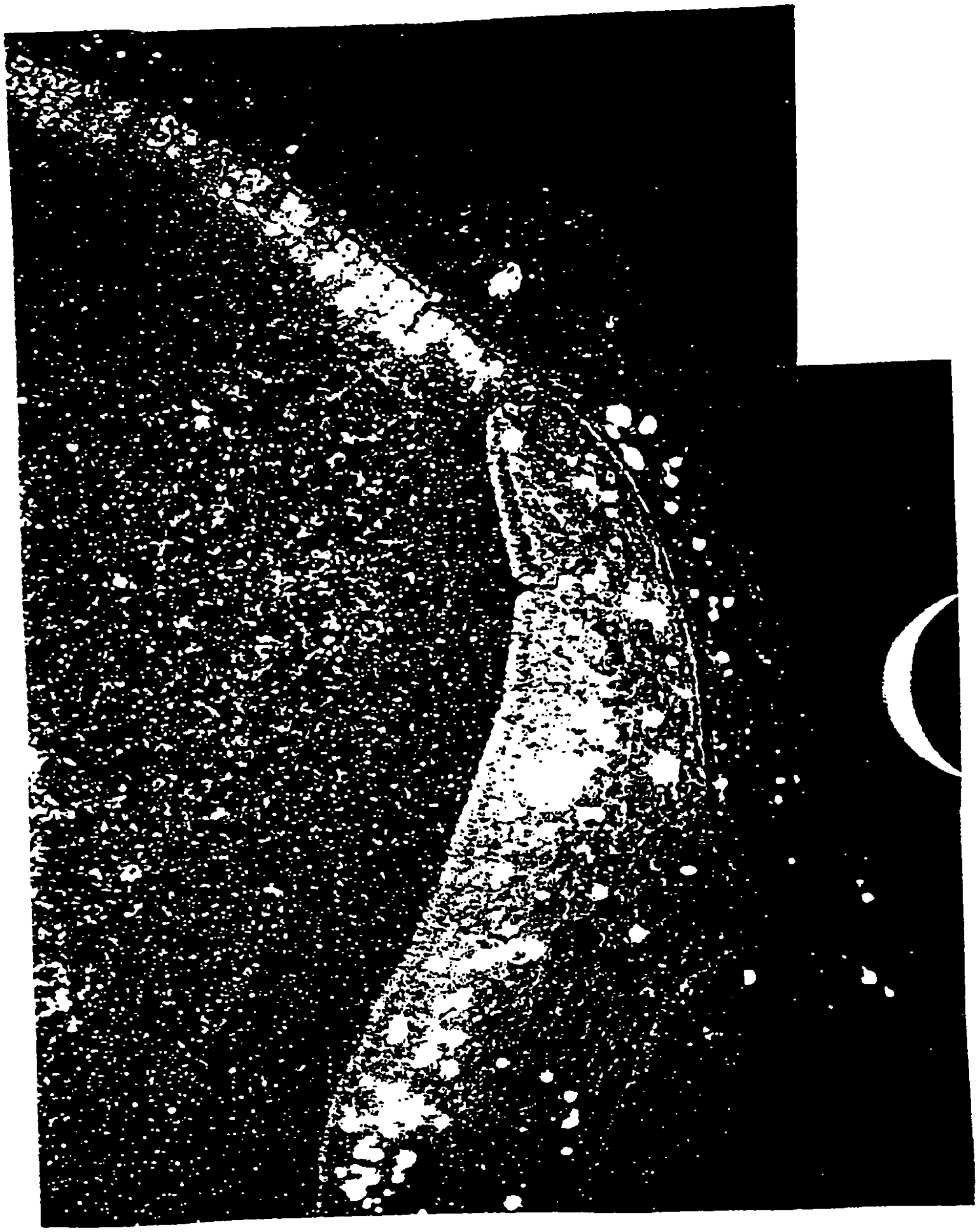


Fig 3

```
   1  GAGCTCCAAGGCATCACCAAGCTTCTATGACGCCAAAACATCCAAGAAAGATATGTACTAGGATACCAAGCACCC

  76  AAGAGTAAACGGAGGAAGTATAATATAAGGCCCTGTTTGATAACAAAGTAGTAAAAAAACTAAAGTATTAAAAAC

 151  TGCAGTAATTTTACGTGTAGATAGAAAATACCATGGTTTTAATATAATAATATTTTTTGCAGTATTCACAATGTA

 226  GAGAAACTGTTTGATTACGCCACATATTACTGCAGTTTAGATCGAGCAAGTACACGGGAAGAAGATAACGACGTC

 301  CCACCCCTTCTTTTCGCCTTCTCTGTTTTTTAAAAAGAGGTCTGGGGTTAGTTTTTTCAATACTGCAGTTTTAAA

 376  ATCACAATTCTTAGAGGCAACCAAACACCTCATTGTAAATAAAACTATGATAATCTCCAAAACTGCAGTATTCTA

 451  AAAATACTACAAAAATTCTTTGTTATCAAACAGGGCCTAAGGAGTTAAAAAAATTTAGCCGTAACTGAGACTCGG

 526  CGAGGCACCAGCAGCTAGCAGTCATCAACACTTGATGGTTGGCAAAGCCGAGTCGACGTGTCGCGGGGCTCGGCC

 601  TGAGCGGGAGATACAATCTGTTCTCCAGTAACCCCGTCGATTTGGCCCGCCGACTAAAGCATCCAGGCATCTCTC

 676  GCTCGAACCCCTATTTAAGCCCCTCCATTCCTCCCAACATTCTCCACACCTCCACGAGTTGCTCATCACTAGCTA

 751  GTACGTTGTACTGTTAGCTACAGATTAAGAAGTGATC ATG GCC CGC GCT CAG GTA CTG CTC ATG
                                            M   A   R   A   Q   V   L   L   M
 815  GCC GCC GCC TTG GTG CTG ATG CTC ACG GCG GCC CCG CGC GCT GCC GTG GCC CTC AAC
       A   A   A   L   V   L   M   L   T   A   A   P   R   A   A   V   A   L   N
 872  TGC GGC CAG GTT GAC AGC AAG ATG AAA CCT TGC CTG ACC TAC GTT CAG GGC GGC CCC
       C   G   Q   V   D   S   K   M   K   P   C   L   T   Y   V   Q   G   G   P
 929  GGC CCG TCC GGC GAA TGC TGC AAC GGC GTC AGG GAT CTC CAT AAC CAG GCG CAA TCC
       G   P   S   G   E   C   C   N   G   V   R   D   L   H   N   Q   A   Q   S
 986  TCG GGC GAC CGC CAA ACC GTT TGC AAC TGC CTG AAG GGG ATC GCT CGC GGC ATC CAC
       S   G   D   R   Q   T   V   C   N   C   L   K   G   I   A   R   G   I   H
1043  AAT CTC AAC CTC AAC AAC GCC GCC AGC ATC CCC TCC AAG TGC AAT GTC AAC GTC CCA
       N   L   N   L   N   N   A   A   S   I   P   S   K   C   N   V   N   V   P
1100  TAC ACC ATC AGC CCC GAC ATC GAC TGC TCC AG gtgattaaatttacactcatccagagtgaaat
       Y   T   I   S   P   D   I   D   C   S   R
1164  ctttaaaaagaactatatttacgaacggagtgagtatataggaacattcatccacgtaaaatttgttgatattaa

1240  cattaacacgcatgattgacctgcag G ATT TAC TGAGCGACGATCCGTCAAGCTGGTGCTCAGCTCATCGA
                                    I   Y   *
1310  TCCACGTGGAGCTGAAGCGCGCAGCCTCTGTCCCTATGTAGTATGGCTACCAGTTATGCCGAGTTTATGCTGAAT

1385  AAGAACTCTCTCCTGTACTCCTTTGGAGGAGATCAGTATCTATGTACGTGAGAGTTGAGAGTTTGTACCATCGGC

1460  ACTCCCAGTGTTTATGGACTATATGCAT
```

Fig4

GTCCACAACTCATGAGCATCACGGAATGGCATGAGTTGAAATATAACTACATTGCTCAAA -1621
GCAACAAAAAGCACATTAGAATCTTGAGCATTGAGATAAGAGTTTTTCTCATGCTCTAAA -1561
TATATATTTTGAGAATCCTTTGGAGGAGAAAAATCCATATTTACAATTCGTTGTAAATTT -1501
GAGTCCATGATCCTAAAGAGATTAAGCATGCGAATTACCCAAACATCAAAATTTGTGCCA -1441
TTGAAACTAAGAGTGTTAGAGAATCCTAATCCCCTAGTTGACATACTTACTCTCTAGGTG -1381
GTGAAACCTAATAATGAGAGATCTAGCTCTAATACCAATTGAGAGGATGTGGATGTCGCC -1321
TAGAGGGGCGGTGAATAGGCGCTTTAAAATAATTACGGTTTAGGCTCGAACAAATGTGGA -1261
ATAAAACTAACGTTTCATTTGTCAAGCGCAAAACCTAAAACAACTAGGCTCACCTATGTG -1201
CACCAACAACTTATGATAAGCAAGATAAAAAAACTAAGTGATGGCAGAATATATAACAAG -1141
AAACAATATGGCTATCACAAAGTGAAGTGCATAAGTAAACAGCTCGGGTAAGGGACAACC -1081
GAGCCATGCGGAGACGACGATGTATCCTCAAGTTCACACACTTGCGGATGCTAATCTCCG -1021
TTTGAAGCAGTGTGGAGGCACAATCGTCCCCAAGAAGCCACTAAGGCCACCGTAATCTCC -961
TCACGCCCTCGCACAATCGAAGATGTTGTGATTCCACTAAGGGACCCTTGAGGGCAGTCA -901
CTGAACCCGTATAAACATGGTTGGAACAATCTCCACGACTTAATTGGAGACTCCCAACAA -841
CACCACGAACCTTCATCATAACGAAATATGGCTTCGAGGTAACCTCAAATGCTCGGGGCA -781
ATTTTTACAACCTAATTGAAGACCTCGACGCTTGCGTGGAGCTTTACACTATAATGATTG -721
AGCTCCAAGGGCATCACCAAGCTTCTATGACGCCAAAACATCCAAGAAAGATATGTACTA -661
GGATACCAAGCACCCAAGAGTAAACGGAGGAAGTATAATATAAGGCCCTGTTTGATAACA -601
AAGTAGTAAAAAAACTAAAGTATTAAAAACTGCAGTAATTTTACGTGTAGATAGAAAATA -541
CCATGGTTTTAATATAATAATATTTTTTGCAGTATTCACAATGTAGAGAAACTGTTTGAT -481
TACGCCACATATTACTGCAGTTTAGATCGAGCAAGTACACGGGAAGAAGATAACGACGTC -421
CCACCCCTTCTTTTCGCCTTCTCTGTTTTTTAAAAAGAGGTCTGGGGTTAGTTTTTTCAA -361
TACTGCAGTTTTAAAATCACAATTCTTAGAGGCAACCAAACACCTCATTGTAAATAAAAC -301
TATGATAATCTCCAAAACTGCAGTATTCTAAAAATACTACAAAAATTCTTTGTTATCAAA -241
CAGGGCCTAAGGAGTTAAAAAAATTTAGCCGTAACTGAGACTCGGCGAGGCACCAGCAGC -181

Fig 4 contd.

```
TAGCAGTCATCAACACTTGATGGTTGGCAAAGGCGAGTCGACGTGTCGCGGGGCTCGGCC  -121

TGAGCGGGAGATACAATCTGTTCTCCAGTAACCCCGTCGATTTGGCCCGCCGACTAAAGC   -61

ATCCAGGCATCTCTCGCTCGAACCCCTATTTAAGCCCCTCCATTCCTCCCAACATTCTCC    -1

ACACCTCCACGAGTTGCTCATCACTAGCTAGTACGTTGTACTGTTAGCTACAGATTAAGA    60

AGTGATCATGGCCCGCGCTCAGGTACTGCTCATGGCCGCCGCCTTGGTGCTGATGCTCAC   120
          M  A  R  A  Q  V  L  L  M  A  A  A  L  V  L  M  L  T

GGCGGCCCCGCGCGCTGCCGTGGCCCTCAACTGCGGCCAGGTTGACAGCAAGATGAAACC   180
  A  A  P  R  A  A  V  A  L  N  C  G  Q  V  D  S  K  M  K  P

TTGCCTGACCTACGTTCAGGGCGGCCCCGGCCCGTCCGGCGAATGCTGCAACGGCGTCAG   240
  C  L  T  Y  V  Q  G  G  P  G  P  S  G  E  C  C  N  G  V  R

GGATCTCCATAACCAGGCGCAATCCTCGGGCGACCGCCAAACCGTTTGCAACTGCCTGAA   300
  D  L  H  N  Q  A  Q  S  S  G  D  R  Q  T  V  C  N  C  L  K

GGGGATCGCTCGCGGCATCCACAATCTCAACCTCAACAACGCCGCCAGCATCCCCTCCAA   360
  G  I  A  R  G  I  H  N  L  N  L  N  N  A  A  S  I  P  S  K

GTGCAATGTCAACGTCCCATACACCATCAGCCCCGACATCGACTGCTCCAGgtgattaaa   420
  C  N  V  N  V  P  Y  T  I  S  P  D  I  D  C  S  R

tttacactcatccagagtgaaatctttaaaaagaactatatttacgaacggagtgagtat   480

ataggaacattcatccacgtaaaatttgttgatattaacattaacacgcatgattgacct   540

gcagGATTTACTGAGCGACGATCCGTCAAGCTGGTGCTCAGCTCATCGATCCACGTGGAG   600
      I  Y  *

CTGAAGCGCGCAGCCTCTGTCCCTATGTAGTATGGCTACCAGTTATGCCGAGTTTATGCT   660

GAATAAGAACTCTCTCCTGTACTCCTTTGGAGGAGATCAGTATCTATGTACGTGAGAGTT   720

GAGAGTTTGTACCATCGGCACTCCCAGTGTTTATGGACTATATGCATACACCTCCTTCTG   780

TGCTCAGTGTGTAACTTGTCTCTCTGTTTCCTCACGTTCGCGTCTCATATAATAATTTAC   840

TTATGTGCTCTAGGATCGTAGTACAGTATCATATATATACCTCTCTATGAATTAGTTTAC   900

CGTAGACCGTATGTTTCTTGAATCTGGATGAAAATTACGGATTCAAGCGTGCGTCCCGCA   960

TATAATAAGCTTGCTTACGGATTCAAGCGTGCGTCACGCGGCTCAGTAGATGATGAGGAT  1020

ACTCGCTGCTGCATCTCTACATCCCGCTCATGAGCTGAGCTGAGCCCGGGTCCTCCCCCG  1080

CTCCGGCCCGCTGGCCACCCCGGCCGGCCGACCCTCAAACAGCCTTCATGACGAGCCGCC  1140

CGCCAGCAAGATCTGTTGGCTCCTCCCCTGTCCGTCGTAGAGAAACCCAGCA          1192
```

Fig 5

```
1   GAGCTCCAAGGCATCACCAAGCTTCTATGACGCCAAAACATCCAAGAAAGATATGTACTAGGATACCAAGCACCC
76  AAGAGTAAACGGAGGAAGTATAATATAAGGCCCTGTTTGATAACAAAGTAGTAAAAAAACTAAAGTATTAAAAAC
151 TGCAGTAATTTTACGTGTAGATAGAAAATACCATGGTTTTAATATAATAATATTTTTTGCAGTATTCACAATGTA
226 GAGAAACTGTTTGATTACGCCACATATTACTGCAGTTTAGATCGAGCAAGTACACGGGAAGAAGATAACGACGTC
301 CCACCCCTTCTTTTCGCCTTCTCTGTTTTTTAAAAAGAGGTCTGGGGTTAGTTTTTTCAATACTGCAGTTTTAAA
376 ATCACAATTCTTAGAGGCAACCAAACACCTCATTGTAAATAAAACTATGATAATCTCCAAAACTGCAGTATTCTA
451 AAAATACTACAAAAATTCTTTGTTATCAAACAGGGCCTAAGGAGTTAAAAAAATTTAGCCGTAACTGAGACTCGG
526 CGAGGCACCAGCAGCTAGCAGTCATCAACACTTGATGGTTGGCAAAGCCGAGTCGACGTGTCGCGGGGCTCGGCC
601 TGAGCGGGAGATACAATCTGTTCTCCAGTAACCCCGTCGATTTGGCCCGCCGACTAAAGCATCCAGGCATCTCTC
676 GCTCGAACCCCTATTTAAGCCCCTCCATTCCTCCCAACATTCTCCACACCTCCACGAGTTGCTCATCACTAGCTA
751 GTACGTTGTACTGTTAGCTACAGATTAAGAAGTGATC
```

Fig 6

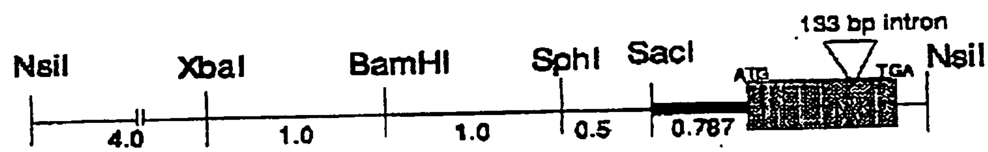

133 bp intron
NsiI
XbaI
BamHI
SphI
SacI
ATG
TGA
NsiI
4.0
1.0
1.0
0.5
0.787

Fig. 7

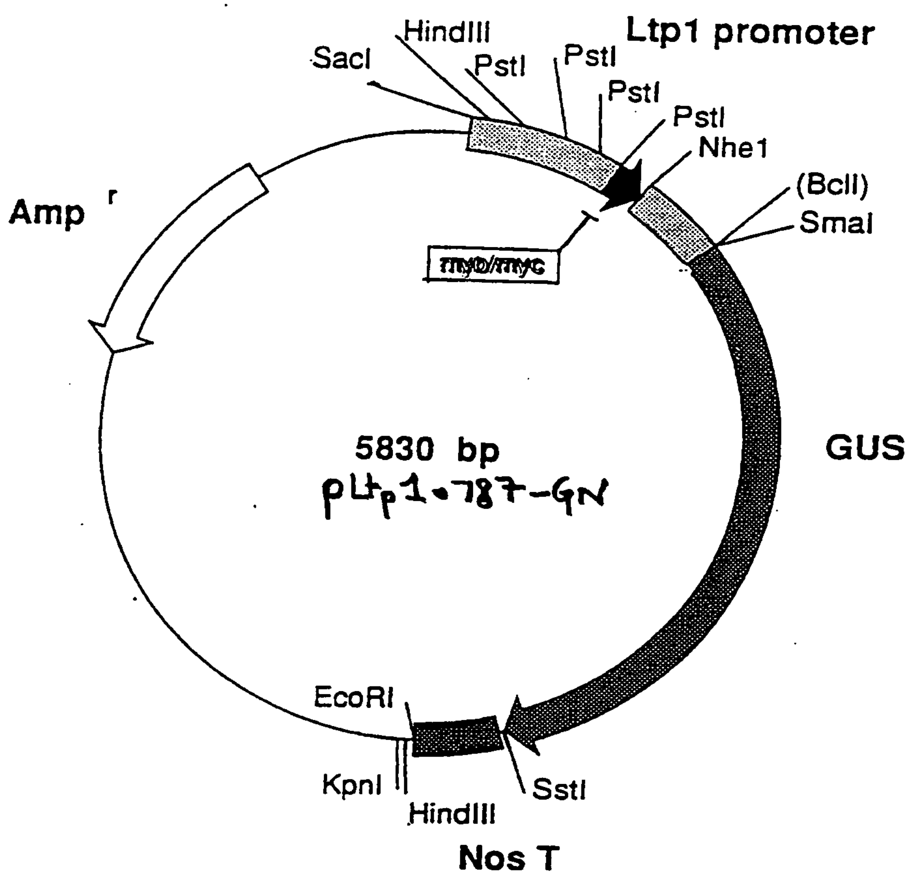
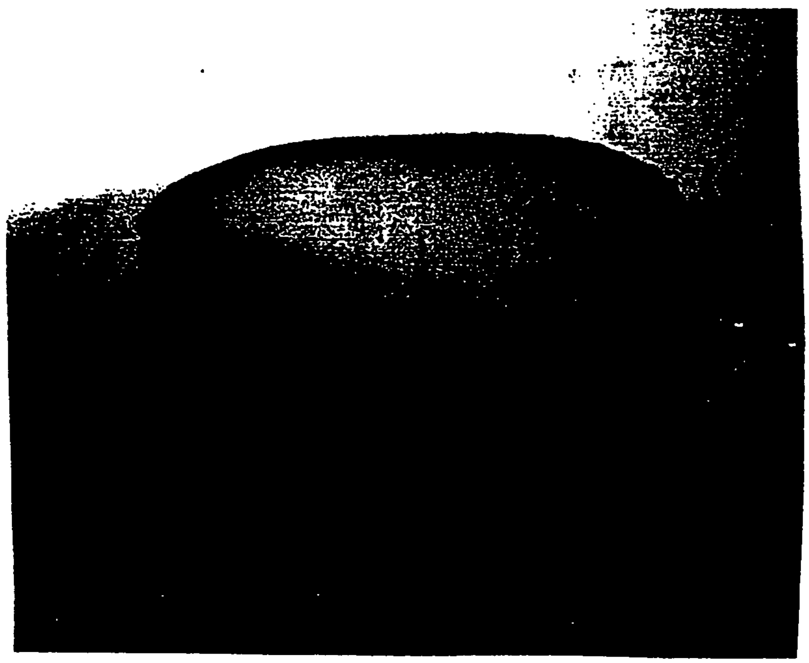

HindIII
SacI
PstI
PstI
PstI
PstI
Nhe1
Ltp1 promoter
(BclI)
SmaI
Amp r
myb/myc
5830 bp
pLtp1.787-GN
GUS
EcoRI
KpnI
HindIII
SstI
Nos T

Fig. 8

Fig. 9



INTERNATIONAL SEARCH REPORT

Inter 'onal Application No
PC:/NO 95/00042

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/82 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
O,X	J. CELL. BIOCHEM. SUPPL. 0, 1994 page 99 O.-A. OLSEN ET AL.; 'The barley LTP2 promoter yields high level of GUS expression in the aleurone layer of developing grains of transgenic rice' see abstract no. X1-213. & Keystone Symposium on improved crop and plant products through biotechnology, Keystone, Colorado, USA, January 9-16, 1994. --- -/--	1-21, 25-30, 44-47

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 June 1995	17.07.95

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Yeats, S

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PLANT MOL. BIOL., vol. 18, no. 19, 1992 page 585-589 K. SKRIVER ET AL.; 'Structure and expression of the barley lipid transfer protein gene Ltp1' cited in the application see the whole document.	1-21, 23-30, 44,45,47
X	PLANT J., vol. 2, 1992 pages 855-862, A.J. FLEMING ET AL.; 'Expression pattern of a tobacco lipid transfer protein gene within the shoot apex' cited in the application see the results section.	1-22,47
X	PLANT CELL, vol. 3, 1991 pages 923-933, L. SOSSOUNTZOV ET AL.; 'Spatial and temporal expression of a maize lipid transfer protein gene' see pages 923-925.	1-22, 26-29, 44,47
A	PLANTA, vol. 192, 1994 pages 574-580, K. GAUSING; 'Lipid transfer protein genes specifically expressed in barley leaves and coleoptiles' see the abstract, Figure 2 and page 577.	1
X	PLANT PHYSIOL., vol. 97, 1991 pages 841-843, C. LINNESTAD ET AL.; 'Promoter of a lipid transfer protein gene expressed in barley aleurone cells contains similar myb and myc recognition sites as the maize Bz-MyC allele' cited in the application see the whole document.	21,23, 24,26, 28,44, 45,47

30

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12P 7/00	A2	(11) International Publication Number: WO 99/64616 (43) International Publication Date: 16 December 1999 (16.12.99)

(21) International Application Number: PCT/US99/13332

(22) International Filing Date: 11 June 1999 (11.06.99)

(30) Priority Data:
60/089,149 12 June 1998 (12.06.98) US

(71) Applicant: ABBOTT LABORATORIES [US/US]; CHAD 0377/AP6D-2, 100 Abbott Park Road, Abbott Park, IL 60064-6050 (US).

(72) Inventors: MUKERJI, Pradip; 1069 Arcaro Drive, Gahanna, OH 43230 (US). KNUTZON, Deborah; 6110 Rockhurst Way, Granite Bay, CA 95746 (US).

(74) Agents: BECKER, Cheryl, L. et al.; Abbott Laboratories, CHAD 0377/AP6D-2, 100 Abbott Park Road, Abbott Park, IL 60064-6050 (US).

(81) Designated States: AU, BG, BR, CA, CN, CZ, HU, IL, JP, KR, MX, NO, NZ, PL, RO, SI, SK, TR, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: POLYUNSATURATED FATTY ACIDS IN PLANTS

(57) Abstract

The present invention relates to compositions and methods for preparing polyunsaturated long chain fatty acids in plants, plant parts and plant cells, such as leaves, roots, fruits and seeds. Nucleic acid sequences and constructs encoding fatty acid desaturases, including Δ5-desaturases, Δ6-desaturases and Δ12-desaturases, are used to generate transgenic plants, plant parts and cells which contain and express one or more transgenes encoding one or more desaturases. Expression of the desaturases with different substrate specificities in the plant system permits the large scale production of polyunsaturated long chain fatty acids such as docosahexaenoic acid, eicosapentaenoic acid, α-linolenic acid, gamma-linolenic acid, arachidonic acid and the like for modification of the fatty acid profile of plants, plant parts and tissues. Manipulation of the fatty acid profiles allows for the production of commercial quantities of novel plant oils and products.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

POLYUNSATURATED FATTY ACIDS IN PLANTS

Field of the Invention

This invention relates to modulating levels of enzymes and/or enzyme components capable of altering the production of long chain polyunsaturated fatty acids (PUFAS) in a host plant. The invention is exemplified by the production of PUFAS in plants.

Background

Three main families of polyunsaturated fatty acids (PUFAs) are the 3 fatty acids, exemplified by arachidonic acid, the ω9 fatty acids exemplified by Mead acid, and the ω6 fatty acids, exemplified by eicosapentaenoic acid. PUFAs are important components of the plasma membrane of the cell, where they may be found in such forms as phospholipids. PUFAs also serve as precursors to other molecules of importance in human beings and animals, including the prostacyclins, leukotrienes and prostaglandins. PUFAs are necessary for proper development, particularly in the developing infant brain, and for tissue formation and repair.

Four major long chain PUFAs of importance include docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA), which are primarily found in different types of fish oil, gamma-linolenic acid (GLA), which is found in the seeds of a number of plants, including evening primrose (*Oenothera biennis*), borage (*Borago officinalis*) and black currants (*Ribes nigrum*), and stearidonic acid (SDA), which is found in marine oils and plant seeds. Both GLA and another important long chain PUFA, arachidonic acid (ARA), are found in filamentous fungi. ARA can be purified from animal tissues including liver and adrenal gland. Mead acid accumulates in essential fatty acid deficient animals.

For DHA, a number of sources exist for commercial production including a variety of marine organisms, oils obtained from cold water marine fish, and egg yolk fractions. For ARA, microorganisms including the genera *Mortierella, Entomophthora, Phytium* and *Porphyridium* can be used for commercial production. Commercial sources of SDA include the genera *Trichodesma and*

Echium. Commercial sources of GLA include evening primrose, black currants and borage. However, there are several disadvantages associated with commercial production of PUFAs from natural sources. Natural sources of PUFAs, such as animals and plants, tend to have highly heterogeneous oil compositions. The oils obtained from these sources therefore can require extensive purification to separate out one or more desired PUFAs or to produce an oil which is enriched in one or more PUFA. Natural sources also are subject to uncontrollable fluctuations in availability. Fish stocks may undergo natural variation or may be depleted by overfishing. Fish oils have unpleasant tastes and odors, which may be impossible to economically separate from the desired product, and can render such products unacceptable as food supplements. Animal oils, and particularly fish oils, can accumulate environmental pollutants. Weather and disease can cause fluctuation in yields from both fish and plant sources. Cropland available for production of alternate oil-producing crops is subject to competition from the steady expansion of human populations and the associated increased need for food production on the remaining arable land. Crops which do produce PUFAs, such as borage, have not been adapted to commercial growth and may not perform well in monoculture. Growth of such crops is thus not economically competitive where more profitable and better established crops can be grown. Large scale fermentation of organisms such as *Mortierella* is also expensive. Natural animal tissues contain low amounts of ARA and are difficult to process. Microorganisms such as *Porphyridium* and *Mortierella* are difficult to cultivate on a commercial scale.

Dietary supplements and pharmaceutical formulations containing PUFAs can retain the disadvantages of the PUFA source. Supplements such as fish oil capsules can contain low levels of the particular desired component and thus require large dosages. High dosages result in ingestion of high levels of undesired components, including contaminants. Care must be taken in providing fatty acid supplements, as overaddition may result in suppression of endogenous biosynthetic pathways and lead to competition with other necessary fatty acids in various lipid fractions *in vivo*, leading to undesirable results. For example, Eskimos having a diet high in ω3 fatty acids have an increased tendency to bleed (U.S. Pat. No.

4,874,603). Unpleasant tastes and odors of the supplements can make such regimens undesirable, and may inhibit compliance by the patient.

A number of enzymes are involved in PUFA biosynthesis. Linoleic acid (LA, 18:2 Δ9, 12) is produced from oleic acid (18:1 Δ9) by a Δ12-desaturase. GLA (18:3 Δ6, 9, 12) is produced from linoleic acid (LA, 18:2 Δ9, 12) by a Δ6-desaturase. ARA (20:4 Δ5, 8, 11, 14) production from DGLA (20:3 Δ8, 11, 14) is catalyzed by a Δ5-desaturase. However, animals cannot desaturate beyond the Δ9 position and therefore cannot convert oleic acid (18:1 Δ9) into linoleic acid (18:2 Δ9, 12). Likewise, α-linolenic acid (ALA, 18:3 Δ9, 12, 15) cannot be synthesized by mammals. Other eukaryotes, including fungi and plants, have enzymes which desaturate at positions Δ21 and Δ15. The major poly-unsaturated fatty acids of animals therefore are either derived from diet and/or from desaturation and elongation of linoleic acid (18:2 Δ9, 12) or α-linolenic acid (18:3 Δ9, 12, 15).

Poly-unsaturated fatty acids are considered to be useful for nutritional, pharmaceutical, industrial, and other purposes. An expansive supply of poly-unsaturated fatty acids from natural sources and from chemical synthesis are not sufficient for commercial needs. Therefore it is of interest to obtain genetic material involved in PUFA biosynthesis from species that naturally produce these fatty acids and to express the isolated material alone or in combination in a heterologous system which can be manipulated to allow production of commercial quantities of PUFAS.

SUMMARY OF THE INVENTION

Novel compositions and methods are provided for preparation of poly-unsaturated long chain fatty acids and desaturases in plants and plant cells. The methods involve growing a host plant cell of interest transformed with an expression cassette functional in a host plant cell, the expression cassette comprising a transcriptional and translational initiation regulatory region, joined in reading frame 5' to a DNA sequence encoding a desaturase polypeptide capable of modulating the production of PUFAs. Expression of the desaturase polypeptide provides for an alteration in the PUFA profile of host plant cells as a result of

altered concentrations of enzymes involved in PUFA biosynthesis. Of particular interest is the selective control of PUFA production in plant tissues and/or plant parts such as leaves, roots, fruits and seeds. The invention finds use for example in the large scale production of DHA, Mead Acid EPA, ARA, Stearidonic acid and GLA and for modification of the fatty acid profile of edible plant tissues and/or plant parts.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows possible pathways for the synthesis of Mead acid (20:3 Δ5, 8, 11), arachidonic acid (20:4 Δ5, 8, 11, 14) and stearidonic acid (18:4 Δ6, 9, 12, 15) from palmitic acid (C_{16}) from a variety of organisms, including algae, *Mortierella* and humans. These PUFAs can serve as precursors to other molecules important for humans and other animals, including prostacyclins, leukotrienes, and prostaglandins, some of which are shown.

Figure 2 shows possible pathways for production of PUFAs in addition to ARA, including taxoleic acid and pinolenic, again compiled from a variety of organisms.

BRIEF DESCRIPTION OF THE SEQUENCE LISTINGS

SEQ ID NO:1 shows DNA sequence from a *Schizochytrium* clone with homology to both Δ12 and Δ15 desaturases.

SEQ ID NO 2 shows peptide sequence from a *Schizochytrium* clone with homology to both Δ12 and Δ15 desaturases.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In order to ensure a complete understanding of the invention, the following definitions are provided:

Δ5-Desaturase: Δ5 desaturase is an enzyme which introduces a double bond between carbons 5 and 6 from the carboxyl end of a fatty acid molecule.

Δ6-Desaturase: Δ6-desaturase is an enzyme which introduces a double bond between carbons 6 and 7 from the carboxyl end of a fatty acid molecule.

Δ9-Desaturase: Δ9-desaturase is an enzyme which introduces a double bond between carbons 9 and 10 from the carboxyl end of a fatty acid molecule.

Δ12-Desaturase: Δ12-desaturase is an enzyme which introduces a double bond between carbons 12 and 13 from the carboxyl end of a fatty acid molecule.

Fatty Acids: Fatty acids are a class of compounds containing a long hydrocarbon chain and a terminal carboxylate group. Fatty acids include the following:

Fatty Acid		
12:0	lauric acid	
16:0	Palmitic acid	
16:1	Palmitoleic acid	
18:0	stearic acid	
18:1	oleic acid	Δ9-18:1
18:2 Δ5,9	Taxoleic acid	Δ5,9-18:2
18:2 Δ6,9	6,9-octadecadienoic acid	Δ6,9-18:2
18:2	Linoleic acid	Δ9,12-18:2 (LA)
18:3 Δ6,9,12	Gamma-linolenic acid	Δ6,9,12-18:3 (GLA)
18:3 Δ5,9,12	Pinolenic acid	Δ5,9,12-18:3
18:3	alpha-linolenic acid	Δ9,12,15-18:3 (ALA)
18:4	Stearidonic acid	Δ6,9,12,15-18:4 (SDA)
20:0	Arachidic acid	
20:1	Eicosenic Acid	
20:2 Δ8, 11		Δ8, 11
20:3 Δ5, 8, 11	Mead Acid	Δ5, 8, 11
22:0	Behehic acid	
22:1	erucic acid	
22:2	Docasadienoic acid	
20:4 ω6	arachidonic acid	Δ5,8,11,14-20:4 (ARA)
20:3 ω6	ω6-eicosatrienoic dihomo-gamma linolenic	Δ8,11,14-20:3 (DGLA)
20:5 ω3	Eicosapentanoic (Timnodonic acid)	Δ5,8,11,14,17-20:5 (EPA)
20:3 ω3	ω3-eicosatrienoic	Δ11,16,17-20:3
20:4 ω3	ω3-eicosatetraenoic	Δ8,11,14,17-20:4
22:5 ω3	Docosapentaenoic	Δ7,10,13,16,19-22:5 (ω3DPA)

Fatty Acid		
22:6 ω3	Docosahexaenoic (cervonic acid)	Δ4,7,10,13,16,19-22:6 (DHA)
24:0	Lignoceric acid	

Taking into account these definitions, the present invention is directed to novel DNA sequences, DNA constructs, methods and compositions are provided which permit modification of the poly-unsaturated long chain fatty acid content of plant cells. Plant cells are transformed with an expression cassette comprising a DNA encoding a polypeptide capable of increasing the amount of one or more PUFA in a plant cell. Desirably, integration constructs may be prepared which provide for integration of the expression cassette into the genome of a host cell. Host cells are manipulated to express a sense or antisense DNA encoding a polypeptide(s) that has desaturase activity. By "desaturase" is intended a polypeptide which can desaturate one or more fatty acids to produce a mono- or poly-unsaturated fatty acid or precursor thereof of interest. By "polypeptide" is meant any chain of amino acids, regardless of length or post-translational modification, for example, glycosylation or phosphorylation. The substrate(s) for the expressed enzyme may be produced by the host cell or may be exogenously supplied.

To achieve expression in a host cell, the transformed DNA is operably associated with transcriptional and translational initiation and termination regulatory regions that are functional in the host cell. Constructs comprising the gene to be expressed can provide for integration into the genome of the host cell or can autonomously replicate in the host cell. For production of linoleic acid (LA), the expression cassettes generally used include a cassette which provides for Δ12 desaturase activity, particularly in a host cell which produces or can take up oleic acid. For production of ALA, the expression cassettes generally used include a cassette which provides for Δ15 or ω3 desaturase activity, particularly in a host cell which produces or can take up LA. For production of GLA or SDA, the expression cassettes generally used include a cassette which provides for Δ6 desaturase activity, particularly in a host cell which produces or can take up LA or ALA.

respectively. Production of ω6-type unsaturated fatty acids, such as LA or GLA, is favored in a plant capable of producing ALA by inhibiting the activity of a Δ15 or ω3 type desaturase; this is accomplished by providing an expression cassette for an antisense Δ15 or ω3 transcript, or by disrupting a Δ15 or ω3 desaturase gene. Similarly, production of LA or ALA is favored in a plant having Δ6 desaturase activity by providing an expression cassette for an antisense Δ6 transcript, or by disrupting a Δ6 desaturase gene. Production of oleic acid likewise is favored in a plant having Δ12 desaturase activity by providing an expression cassette for an antisense Δ12 transcript, or by disrupting a Δ12 desaturase gene. For production of ARA, the expression cassette generally used provides for Δ5 desaturase activity, particularly in a host cell which produces or can take up DGLA. Production of ω6-type unsaturated fatty acids, such as ARA, is favored in a plant capable of producing ALA by inhibiting the activity of a Δ15 or ω3 type desaturase; this is accomplished by providing an expression cassette for an antisense Δ15 or ω3 transcript, or by disrupting a Δ15 or ω3 desaturase gene.

TRANSGENIC PLANT PRODUCTION OF FATTY ACIDS

Transgenic plant production of PUFAs offers several advantages over purification from natural sources such as fish or plants. Production of fatty acids from recombinant plants provides the ability to alter the naturally occurring plant fatty acid profile by providing new synthetic pathways in the host or by suppressing undesired pathways, thereby increasing levels of desired PUFAs, or conjugated forms thereof, and decreasing levels of undesired PUFAs. Production of fatty acids in transgenic plants also offers the advantage that expression of desaturase genes in particular tissues and/or plant parts means that greatly increased levels of desired PUFAs in those tissues and/or parts can be achieved, making recovery from those tissues more economical. For example, the desired PUFAs can be expressed in seed; methods of isolating seed oils are well established. In addition to providing a source for purification of desired PUFAs, seed oil components can be manipulated through expression of desaturase genes, either alone or in combination with other genes such as elongases, to provide seed oils having a particular PUFA profile in concentrated form. The concentrated seed oils then can be added to animal milks

and/or synthetic or semi-synthetic milks to serve as infant formulas where human nursing is impossible or undesired, or in cases of malnourishment or disease in both adults and infants.

For production of PUFAs, depending upon the host cell, the availability of substrate, and the desired end product(s), several polypeptides, particularly desaturases, are of interest including those polypeptides which catalyze the conversion of stearic acid to oleic acid, LA to GLA, of ALA to SDA, of oleic acid to LA, or of LA to ALA, which includes enzymes which desaturate at the Δ6, Δ9, Δ12, Δ15 or ω3 positions. Considerations for choosing a specific polypeptide having desaturase activity include the pH optimum of the polypeptide, whether the polypeptide is a rate limiting enzyme or a component thereof, whether the desaturase used is essential for synthesis of a desired poly-unsaturated fatty acid, and/or co-factors required by the polypeptide. The expressed polypeptide preferably has parameters compatible with the biochemical environment of its location in the host cell. For example, the polypeptide may have to compete for substrate with other enzymes in the host cell. Analyses of the K_m and specific activity of the polypeptide in question therefore are considered in determining the suitability of a given polypeptide for modifying PUFA production in a given host cell. The polypeptide used in a particular situation therefore is one which can function under the conditions present in the intended host cell but otherwise can be any polypeptide having desaturase activity which has the desired characteristic of being capable of modifying the relative production of a desired PUFA. A scheme for the synthesis of arachidonic acid (20:4 Δ5, 8, 11, 14) from palmitic acid (C_{16}) is shown in Figure 1. A key enzyme in this pathway is a Δ5-desaturase which converts DH-γ-linolenic acid (DGLA, eicosatrienoic acid) to ARA. Conversion of α-linolenic acid (ALA) to stearidonic acid by a Δ6-desaturase is also shown. Production of PUFAs in addition to ARA, including EPA and DHA is shown in Figure 2. A key enzyme in the synthesis of arachidonic acid (20:4 Δ5, 8, 11, 14) from stearic acid (C_{18}) is a Δ6-desaturase which converts the linoleic acid into γ-linolenic acid. Conversion of α-linolenic acid (ALA) to stearidonic acid by a Δ6-desaturase also is shown. For production of ARA, the DNA sequence used encodes

a polypeptide having Δ5 desaturase activity. In particular instances, this can be coupled with an expression cassette which provides for production of a polypeptide having Δ6 desaturase activity and, optionally, a transcription cassette providing for production of antisense sequences to a Δ 15 transcription product. The choice of combination of cassettes used depends in part on the PUFA profile of the host cell. Where the host cell Δ5-desaturase activity is limiting, overexpression of Δ5 desaturase alone generally will be sufficient to provide for enhanced ARA production.

SOURCES OF POLYPEPTIDES HAVING DESATURASE ACTIVITY

As sources of polypeptides having desaturase activity and oligonucleotides encoding such polypeptides are organisms which produce a desired polyunsaturated fatty acid. As an example, microorganisms having an ability to produce ARA can be used as a source of Δ5-desaturase genes; microorganisms which GLA or SDA can be used as a source of Δ6-desaturase and/or Δ12-desaturase genes. Such microorganisms include, for example, those belonging to the genera *Mortierella*, *Conidiobolus*, *Pythium*, *Phytophathora*, *Penicillium*, *Porphyridium*, *Coidosporium*, *Mucor*, *Fusarium*, *Aspergillus*, *Rhodotorula*, and *Entomophthora*. Within the genus *Porphyridium*, of particular interest is *Porphyridium cruentum*. Within the genus *Mortierella*, of particular interest are *Mortierella elongata*, *Mortierella exigua*, *Mortierella hygrophila*, *Mortierella ramanniana*, var. angulispora, and *Mortierella alpina*. Within the genus *Mucor*, of particular interest are *Mucor circinelloides* and *Mucor javanicus*.

DNAs encoding desired desaturases can be identified in a variety of ways. As an example, a source of the desired desaturase, for example genomic or cDNA libraries from *Mortierella*, is screened with detectable enzymatically- or chemically-synthesized probes, which can be made from DNA, RNA, or non-naturally occurring nucleotides, or mixtures thereof. Probes may be enzymatically synthesized from DNAs of known desaturases for normal or reduced-stringency hybridization methods. Oligonucleotide probes also can be used to screen sources and can be based on sequences of known desaturases, including sequences

conserved among known desaturases, or on peptide sequences obtained from the desired purified protein. Oligonucleotide probes based on amino acid sequences can be degenerate to encompass the degeneracy of the genetic code, or can be biased in favor of the preferred codons of the source organism. Oligonucleotides also can be used as primers for PCR from reverse transcribed mRNA from a known or suspected source; the PCR product can be the full length cDNA or can be used to generate a probe to obtain the desired full length cDNA. Alternatively, a desired protein can be entirely sequenced and total synthesis of a DNA encoding that polypeptide performed.

Once the desired genomic or cDNA has been isolated, it can be sequenced by known methods. It is recognized in the art that such methods are subject to errors, such that multiple sequencing of the same region is routine and is still expected to lead to measurable rates of mistakes in the resulting deduced sequence, particularly in regions having repeated domains, extensive secondary structure, or unusual base compositions, such as regions with high GC base content. When discrepancies arise, resequencing can be done and can employ special methods. Special methods can include altering sequencing conditions by using: different temperatures; different enzymes; proteins which alter the ability of oligonucleotides to form higher order structures; altered nucleotides such as ITP or methylated dGTP; different gel compositions, for example adding formamide; different primers or primers located at different distances from the problem region; or different templates such as single stranded DNAs. Sequencing of mRNA can also be employed.

For the most part, some or all of the coding sequence for the polypeptide having desaturase activity is from a natural source. In some situations, however, it is desirable to modify all or a portion of the codons, for example, to enhance expression, by employing host preferred codons. Host preferred codons can be determined from the codons of highest frequency in the proteins expressed in the largest amount in a particular host species of interest. Thus, the coding sequence for a polypeptide having desaturase activity can be synthesized in whole or in part. All or portions of the DNA also can be synthesized to remove any destabilizing sequences or regions of secondary structure which would be present in the

transcribed mRNA. All or portions of the DNA also can be synthesized to alter the base composition to one more preferable in the desired host cell. Methods for synthesizing sequences and bringing sequences together are well established in the literature. *In vitro* mutagenesis and selection, site-directed mutagenesis, or other means can be employed to obtain mutations of naturally occurring desaturase genes to produce a polypeptide having desaturase activity *in vivo* with more desirable physical and kinetic parameters for function in the host cell, such as a longer half-life or a higher rate of production of a desired polyunsaturated fatty acid.

Desirable cDNAs have less than 60% A+T composition, preferably less than 50% A+T composition. On a localized scale of a sliding window of 20 base pairs, it is preferable that there are no localized regions of the cDNA with greater than 75% A+T composition; with a window of 60 base pairs, it is preferable that there are no localized regions of the cDNA with greater than 60%, more preferably no localized regions with greater than 55% A+T composition.

Mortierella alpina Desaturases

Of particular interest are the *Mortierella alpina* Δ5-desaturase, Δ6-desaturase, Δ12-desaturase and Δ15 desaturase. The gene encoding the *Mortierella alpina* Δ5-desaturase can be expressed in transgenic plants to effect greater synthesis of ARA from DGLA. Other DNAs which are substantially identical in sequence to the *Mortierella alpina* Δ5-desaturase DNA, or which encode polypeptides which are substantially identical in sequence to the *Mortierella alpina* Δ5-desaturase polypeptide, also can be used. The gene encoding the *Mortierella alpina* Δ6-desaturase can be expressed in transgenic plants or animals to effect greater synthesis of GLA from linoleic acid or of stearidonic acid (SDA) from ALA. Other DNAs which are substantially identical in sequence to the *Mortierella alpina* Δ6-desaturase DNA, or which encode polypeptides which are substantially identical in sequence to the *Mortierella alpina* Δ6-desaturase polypeptide, also can be used.

The gene encoding the *Mortierella alpina* Δ12-desaturase can be expressed in transgenic plants to effect greater synthesis of LA from oleic acid. Other DNAs

which are substantially identical to the *Mortierella alpina* Δ12-desaturase DNA, or which encode polypeptides which are substantially identical to the *Mortierella alpina* Δ12-desaturase polypeptide, also can be used.

By substantially identical in sequence is intended an amino acid sequence or nucleic acid sequence exhibiting in order of increasing preference at least 60%, 80%, 90% or 95% homology to the *Mortierella alpina* Δ5-desaturase amino acid sequence or nucleic acid sequence encoding the amino acid sequence. For polypeptides, the length of comparison sequences generally is at least 16 amino acids, preferably at least 20 amino acids, or most preferably 35 amino acids. For nucleic acids, the length of comparison sequences generally is at least 50 nucleotides, preferably at least 60 nucleotides, and more preferably at least 75 nucleotides, and most preferably, 110 nucleotides. Homology typically is measured using sequence analysis software, for example, the Sequence Analysis software package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wisconsin 53705, MEGAlign (DNAStar, Inc., 1228 S. Park St., Madison, Wisconsin 53715), and MacVector (Oxford Molecular Group, 2105 S. Bascom Avenue, Suite 200, Campbell, California 95008). Such software matches similar sequences by assigning degrees of homology to various substitutions, deletions, and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine and leucine; aspartic acid, glutamic acid, asparagine, and glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine. Substitutions may also be made on the basis of conserved hydrophobicity or hydrophilicity (Kyte and Doolittle, *J. Mol. Biol.* 157: 105-132, 1982), or on the basis of the ability to assume similar polypeptide secondary structure (Chou and Fasman, *Adv. Enzymol.* 47: 45-148, 1978).

EXPRESSION OF DESATURASE GENES

Once the DNA encoding a desaturase polypeptide has been obtained, it is placed in a vector capable of replication in a host cell, or is propagated *in vitro* by means of techniques such as PCR or long PCR. Replicating vectors can include plasmids, phage, viruses, cosmids and the like. Desirable vectors include those

useful for mutagenesis of the gene of interest or for expression of the gene of interest in host cells. The technique of long PCR has made *in vitro* propagation of large constructs possible, so that modifications to the gene of interest, such as mutagenesis or addition of expression signals, and propagation of the resulting constructs can occur entirely *in vitro* without the use of a replicating vector or a host cell.

For expression of a desaturase polypeptide, functional transcriptional and translational initiation and termination regions are operably linked to the DNA encoding the desaturase polypeptide. Transcriptional and translational initiation and termination regions are derived from a variety of nonexclusive sources, including the DNA to be expressed, genes known or suspected to be capable of expression in the desired system, expression vectors, chemical synthesis, or from an endogenous locus in a host cell. Expression in a plant tissue and/or plant part presents certain efficiencies, particularly where the tissue or part is one which is easily harvested, such as seed, leaves, fruits, flowers, roots, etc. Expression can be targeted to that location within the plant by using specific regulatory sequences, such as those of USPN 5,463,174, USPN 4,943,674, USPN 5,106,739, USPN 5,175,095, USPN 5,420,034, USPN 5,188,958, and USPN 5,589,379. Alternatively, the expressed protein can be an enzyme which produces a product which may be incorporated, either directly or upon further modifications, into a fluid fraction from the host plant. In the present case, expression of desaturase genes, or antisense desaturase transcripts, can alter the levels of specific PUFAs, or derivatives thereof, found in plant parts and/or plant tissues. The Δ5-desaturase polypeptide coding region is expressed either by itself or with other genes, in order to produce tissues and/or plant parts containing higher proportions of desired PUFAs or in which the PUFA composition more closely resembles that of human breast milk (Prieto *et al.*, PCT publication WO 95/24494). The termination region can be derived from the 3' region of the gene from which the initiation region was obtained or from a different gene. A large number of termination regions are known to and have been found to be satisfactory in a variety of hosts from the same and different genera and species. The termination region usually is selected more as a matter of convenience rather than because of any particular property.

The choice of a host cell is influenced in part by the desired PUFA profile of the transgenic cell, and the native profile of the host cell. As an example, for production of linoleic acid from oleic acid, the DNA sequence used encodes a polypeptide having Δ12 desaturase activity, and for production of GLA from linoleic acid, the DNA sequence used encodes a polypeptide having Δ6 desaturase activity. Use of a host cell which expresses Δ12 desaturase activity and lacks or is depleted in Δ15 desaturase activity, can be used with an expression cassette which provides for overexpression of Δ6 desaturase alone generally is sufficient to provide for enhanced GLA production in the transgenic cell. Where the host cell expresses Δ9 desaturase activity, expression of both a Δ12- and a Δ6-desaturase can provide for enhanced GLA production. In particular instances where expression of Δ6 desaturase activity is coupled with expression of Δ12 desaturase activity, it is desirable that the host cell naturally have, or be mutated to have, low Δ15 desaturase activity. Alternatively, a host cell for Δ6 desaturase expression may have, or be mutated to have, high Δ12 desaturase activity.

Expression in a host cell can be accomplished in a transient or stable fashion. Transient expression can occur from introduced constructs which contain expression signals functional in the host cell, but which constructs do not replicate and rarely integrate in the host cell, or where the host cell is not proliferating. Transient expression also can be accomplished by inducing the activity of a regulatable promoter operably linked to the gene of interest, although such inducible systems frequently exhibit a low basal level of expression. Stable expression can be achieved by introduction of a construct that can integrate into the host genome or that autonomously replicates in the host cell. Stable expression of the gene of interest can be selected for through the use of a selectable marker located on or transfected with the expression construct, followed by selection for cells expressing the marker. When stable expression results from integration, integration of constructs can occur randomly within the host genome or can be targeted through the use of constructs containing regions of homology with the host genome sufficient to target recombination with the host locus. Where constructs

are targeted to an endogenous locus, all or some of the transcriptional and translational regulatory regions can be provided by the endogenous locus.

When increased expression of the desaturase polypeptide in the source plant is desired, several methods can be employed. Additional genes encoding the desaturase polypeptide can be introduced into the host organism. Expression from the native desaturase locus also can be increased through homologous recombination, for example by inserting a stronger promoter into the host genome to cause increased expression, by removing destabilizing sequences from either the mRNA or the encoded protein by deleting that information from the host genome, or by adding stabilizing sequences to the mRNA (*see* USPN 4,910,141 and USPN 5,500,365.)

When it is desirable to express more than one different gene, appropriate regulatory regions and expression methods, introduced genes can be propagated in the host cell through use of replicating vectors or by integration into the host genome. Where two or more genes are expressed from separate replicating vectors, it is desirable that each vector has a different means of replication. Each introduced construct, whether integrated or not, should have a different means of selection and should lack homology to the other constructs to maintain stable expression and prevent reassortment of elements among constructs. Judicious choices of regulatory regions, selection means and method of propagation of the introduced construct can be experimentally determined so that all introduced genes are expressed at the necessary levels to provide for synthesis of the desired products.

Constructs comprising the gene of interest may be introduced into a host cell by standard techniques. These techniques include transfection, infection, bolistic impact, electroporation, microinjection, scraping, or any other method which introduces the gene of interest into the host cell (*see* USPN 4,743,548, USPN 4,795,855, USPN 5,068,193, USPN 5,188,958, USPN 5,463,174, USPN 5,565,346 and USPN 5,565,347). For convenience, a host cell which has been manipulated by any method to take up a DNA sequence or construct will be referred to as "transformed" or "recombinant" herein. The subject host will have at least have one copy of the expression construct and may have two or more, depending upon

whether the gene is integrated into the genome, amplified, or is present on an extrachromosomal element having multiple copy numbers.

The transformed host cell can be identified by selection for a marker contained on the introduced construct. Alternatively, a separate marker construct may be introduced with the desired construct, as many transformation techniques introduce many DNA molecules into host cells. Typically, transformed hosts are selected for their ability to grow on selective media. Selective media may incorporate an antibiotic or lack a factor necessary for growth of the untransformed host, such as a nutrient or growth factor. An introduced marker gene therefor may confer antibiotic resistance, or encode an essential growth factor or enzyme, and permit growth on selective media when expressed in the transformed host cell. Desirably, resistance to kanamycin and the amino glycoside G418 are of interest (*see* USPN 5,034,322). Selection of a transformed host can also occur when the expressed marker protein can be detected, either directly or indirectly. The marker protein may be expressed alone or as a fusion to another protein. The marker protein can be detected by its enzymatic activity; for example β galactosidase can convert the substrate X-gal to a colored product, and luciferase can convert luciferin to a light-emitting product. The marker protein can be detected by its light-producing or modifying characteristics; for example, the green fluorescent protein of *Aequorea victoria* fluoresces when illuminated with blue light. Antibodies can be used to detect the marker protein or a molecular tag on, for example, a protein of interest. Cells expressing the marker protein or tag can be selected, for example, visually, or by techniques such as FACS or panning using antibodies.

The PUFAs produced using the subject methods and compositions may be found in the host plant tissue and/or plant part as free fatty acids or in conjugated forms such as acylglycerols, phospholipids, sulfolipids or glycolipids, and may be extracted from the host cell through a variety of means well-known in the art. Such means may include extraction with organic solvents, sonication, supercritical fluid extraction using for example carbon dioxide, and physical means such as presses, or combinations thereof. Of particular interest is extraction with hexane or methanol and chloroform. Where desirable, the aqueous layer can be acidified to

protonate negatively charged moieties and thereby increase partitioning of desired products into the organic layer. After extraction, the organic solvents can be removed by evaporation under a stream of nitrogen. When isolated in conjugated forms, the products are enzymatically or chemically cleaved to release the free fatty acid or a less complex conjugate of interest, and are then subjected to further manipulations to produce a desired end product. Desirably, conjugated forms of fatty acids are cleaved with potassium hydroxide.

PURIFICATION OF FATTY ACIDS

If further purification is necessary, standard methods can be employed. Such methods include extraction, treatment with urea, fractional crystallization, HPLC, fractional distillation, silica gel chromatography, high speed centrifugation or distillation, or combinations of these techniques. Protection of reactive groups, such as the acid or alkenyl groups, may be done at any step through known techniques, for example alkylation or iodination. Methods used include methylation of the fatty acids to produce methyl esters. Similarly, protecting groups may be removed at any step. Desirably, purification of fractions containing ARA, DHA and EPA is accomplished by treatment with urea and/or fractional distillation.

USES OF FATTY ACIDS

The uses of the fatty acids of subject invention are several. Probes based on the DNAs of the present invention may find use in methods for isolating related molecules or in methods to detect organisms expressing desaturases. When used as probes, the DNAs or oligonucleotides need to be detectable. This is usually accomplished by attaching a label either at an internal site, for example via incorporation of a modified residue, or at the 5' or 3' terminus. Such labels can be directly detectable, can bind to a secondary molecule that is detectably labeled, or can bind to an unlabelled secondary molecule and a detectably labeled tertiary molecule; this process can be extended as long as is practical to achieve a satisfactorily detectable signal without unacceptable levels of background signal. Secondary, tertiary, or bridging systems can include use of antibodies directed against any other molecule, including labels or other antibodies, or can involve any

molecules which bind to each other, for example a biotin-streptavidin/avidin system. Detectable labels typically include radioactive isotopes, molecules which chemically or enzymatically produce or alter light, enzymes which produce detectable reaction products, magnetic molecules, fluorescent molecules or molecules whose fluorescence or light-emitting characteristics change upon binding. Examples of labelling methods can be found in USPN 5,011,770. Alternatively, the binding of target molecules can be directly detected by measuring the change in heat of solution on binding of probe to target via isothermal titration calorimetry, or by coating the probe or target on a surface and detecting the change in scattering of light from the surface produced by binding of target or probe, respectively, as may be done with the BIAcore system.

The invention will be better understood by reference to the following non-limiting examples.

Examples

Example 1

Expression of ω-3 desaturase from *C. elegans* in transgenic plants.

The D15/ω-3 activity of *Brassica napus* can be increased by the expression of an ω-3 desaturase from *C. elegans*. The fat-1 cDNA clone (Genbank accession L41807; Spychalla, J. P., Kinney, A. J., and Browse, J. 1997 P.N.A.S. 94, 1142-1147 was obtained from John Browse at Washington State University. The fat-1 cDNA was modified by PCR to introduce cloning sites using the following primers:

Fat-1forward:

5'-CUACUACUACUACTGCAGACAATGGTCGCTCATTCCTCAGA-3'

Fat-1reverse:

5'- CAUCAUCAUCAUGCGGCCGCTTACTTGGCCTTTGCCTT - 3'

These primers allowed the amplification of the entire coding region and added PstI and NotI sites to the 5'- and 3'-ends, respectively. The PCR product was subcloned into pAMP1 (GIBCOBRL) using the CloneAmp system (GIBCOBRL)

to create pCGN5562. The sequence was verified by sequencing of both strands to be sure no changes were introduced by PCR. For seed specific expression, the Fat-1 coding region was cut out of pCGN5562 as a PstI/NotI fragment and inserted between the PstI/NotI sites of the binary vector, pCGN8623, to create pCGN5563. PCGN5563 can be introduced into Brassica napus via *Agrobacterium*-mediated transformation.

Construction of pCGN8623

The polylinker region of the napin promoter cassette, pCGN7770, was replaced by ligating the following oligonucleotides:

5'- TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC -3' and

5'- TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3'. These oligonucleotides were ligated into SalI/XhoI-digested pCGN7770 to produce pCGN8619. These oligos encode BamHI, NotI, HindIII, and PstI restriction sites. pCGN8619 contains the oligos oriented such that the PstI site is closest to the napin 5' regulatory region. A fragment containing the napin 5' regulatory region, polylinker, and napin 3' region was removed from pCGN8619 by digestion with Asp718I. The fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the napin promoter was closest to the blunted Asp718I site of pCGN5139 and the napin 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8623.

To produce high levels of stearidonic acid in *Brassica*, the C. elegans ω-3 desaturase can be combined with D6- and D12-desaturases from *Mortierella alpina*. PCGN5563-transformed plants may be crossed with pCGN5544-transformed plants expressing the D6-and D12-desaturases.

The resulting F1 seeds can be analyzed for stearidonic acid content and selected F1 plants can be used for self-pollination to produce F2 seed, or as donors for production of dihaploids, or additional crosses.

An alternative method to combine the fat-1 cDNA with *M. alpina* D6 and D12 desaturases is to combine them on one T-DNA for transformation. The fat-1 coding region from pCGN5562 can be cut out as a PstI/NotI fragment and inserted into PstI/NotI digested pCGN8619. The transcriptional unit consisting of the napin 5' regulatory region, the fat-1 coding region, and the napin 3'-regulatory region can be cut out as a Sse8387I fragment and inserted into pCGN5544 cut with Sse8387I. The resulting plasmid would contain three napin transcriptional units containing the *C. elegans* ω-3 desaturase, *M. alpina* D6 desaturase, and *M. alpina* D12 desaturase, all oriented in the same direction as the 35S/nptII/tml transcriptional unit used for selection of transformed tissue.

Example 2

Over-Expression of D15-desaturase Activity in Transgenic Canola

The D15-desaturase activity of Brassica napus can be increased by over-expression of the D15-desaturase cDNA clone.

A *B. napus* D15-desaturase cDNA clone was obtained by PCR amplification of first-strand cDNA derived from *B. napus* cv. 212/86. The primers were based on published sequence: Genbank # L01418 Arondel et al. 1992 Science 258:1353-1355.

The following primers were used:

Bnd15-FORWARD

5'-CUACUACUACUAGAGCTCAGCGATGGTTGTTGCTATGGAC-3'

Bnd15-REVERSE

5'-CAUCAUCAUCAUGAATTCTTAATTGATTTTAGATTTG-3'

These primers allowed the amplification of the entire coding region and added SacI and EcoRI sites to the 5'- and 3'-ends, respectively

The PCR product was subcloned into pAMP1 (GIBCOBRL) using the CloneAmp system (GIBCOBRL) to create pCGN5520. The sequence was verified by sequencing of both strands to be sure that the open reading frame remained intact.

For seed specific expression, the D15-desaturase coding region was cut out of pCGN5520 as a BamHI/SalI fragment and inserted between the BglII and XhoI sites of the pCGN7770, to create pCGN5557. The PstI fragment of pCGN5557 containing the napin 5'-regulatory region, *B. napus* D15-desaturase, and napin 3'-regulatory region was inserted into the PstI site of the binary vector, pCGN5138 to produce pCGN5558. pCGN5558 was introduced into *Brassica napus* via *Agrobacterium*-mediated transformation.

To produce high levels of stearidonic acid in *Brassica*, the D15-desaturase can be combined with D6- and D12-desaturases from *Mortierella alpina*. PCGN5558-transformed plants may be crossed with pCGN5544-transformed plants expressing the D6 and D12-desaturases. The resulting F1 seeds can be analyzed for stearidonic acid content and selected F1 plants can be used for self-pollination to produce F2 seed, or as donors for production of dihaploids, or additional crosses.

An alternative method to combine the *B. napus* D15-desaturase with *M. alpina* D6 and D12 desaturases is to combine them on one T-DNA for transformation. The transcription cassette consisting of the napin 5'-regulatory region, the D15-desaturase coding region, and the napin 3'-regulatory region can be cut out of pCGN5557 as a SwaI fragment and inserted into SwaI-digested pCGN5544. The resulting plasmid would contain three napin transcriptional units containing the *M. alpina* D6 desaturase, the *B. napus* D15-desaturase, and the *M. alpina* D12 desaturase, all oriented in the same direction as the 35S/nptII/tml transcriptional unit used for selection of transformed tissue.

Example 3

Expression of Δ5 Desaturase in Plants

Expression in Leaves

Ma29 is a putative *M. alpina* D5 desaturase as determined by sequence homology. This experiment was designed to determine whether leaves expressing Ma29 (as determined by Northern) were able to convert exogenously applied DGLA (20:3) to ARA (20:4).

The Ma29 desaturase cDNA was modified by PCR to introduce convenient restriction sites for cloning. The desaturase coding region has been inserted into a d35 cassette under the control of the double 35S promoter for expression in *Brassica* leaves (pCGN5525) following standard protocols (*see* USPN 5,424,200 and USPN 5,106,739). Transgenic *Brassica* plants containing pCGN5525 were generated following standard protocols (*see* USPN 5,188,958 and USPN 5,463,174).

In the first experiment, three plants were used: a control, LPOO4-1, and two transgenics,, 5525-23 and 5525-29. LP004 is a low-linolenic *Brassica* variety. Leaves of each were selected for one of three treatments: water, GLA or DGLA. GLA and DGLA were purchased as sodium salts from NuChek Prep and dissolved in water at 1 mg/ml. Aliquots were capped under N_2 and stored at -70 degrees C. Leaves were treated by applying a 50 µl drop to the upper surface and gently spreading with a gloved finger to cover the entire surface. Applications were made approximately 30 minutes before the end of the light cycle to minimize any photo-oxidation of the applied fatty acids. After 6 days of treatment one leaf from each treatment was harvested and cut in half through the mid rib. One half was washed with water to attempt to remove unincorporated fatty acid. Leaf samples were lyophilized overnight, and fatty acid composition determined by gas chromatography (GC). The results are shown in Table 1.

Table 1

Fatty Acid Analysis of Leaves from Ma29 Transgenic *Brassica* Plants

Treatment	SPL	16:00	16:01	18:00	18:01	18:1o	18:1v	18:02	18:3g	18:03	18:04	20:00	20:01
	#	%	%	%	%	%	%	%	%	%	%	%	%
Water	33	12.95	0.08	2.63	2.51	1.54	0.98	16.76	0	45.52	0	0.09	0
	34	13.00	0.09	2.67	2.56	1.55	1.00	16.86	0	44.59	0	0.15	0
	35	14.13	0.09	2.37	2.15	1.27	0.87	16.71	0	49.91	0	0.05	0.01
	36	13.92	0.08	2.32	2.07	1.21	0.86	16.16	0	50.25	0	0.05	0
	37	13.79	0.11	2.10	2.12	1.26	0.86	15.90	0.08	46.29	0	0.54	0.01
	38	12.80	0.09	1.94	2.08	1.35	0.73	14.54	0.11	45.61	0	0.49	0.01
GLA	39	12.10	0.09	2.37	2.10	1.29	0.82	14.85	1.63	43.66	0	0.53	0
	40	12.78	0.10	2.34	2.22	1.36	0.86	15.29	1.72	47.22	0	0.50	0.02
	41	13.71	0.07	2.68	2.16	1.34	0.82	15.92	2.12	46.55	0	0.09	0
	42	14.10	0.07	2.75	2.35	1.51	0.84	16.66	1.56	46.41	0	0.09	0.01
	43	13.62	0.09	2.22	1.94	1.21	0.73	14.68	2.42	46.69	0	0.51	0.01
	44	13.92	0.09	2.20	2.17	1.32	0.85	15.22	2.30	46.05	0	0.53	0.02
DGLA	45	12.45	0.14	2.30	2.28	1.37	0.91	15.65	0.07	44.62	0	0.12	0.01
	46	12.67	0.15	2.69	2.50	1.58	0.92	15.96	0.09	42.77	0	0.56	0.01
	47	12.56	0.23	3.40	1.98	1.13	0.86	13.57	0.03	45.52	0	0.51	0.01
	48	13.07	0.24	3.60	2.51	1.63	0.88	13.54	0.04	45.13	0	0.50	0.01
	49	13.26	0.07	2.81	2.34	1.67	0.67	16.04	0.04	43.89	0	0.59	0
	50	13.53	0.07	2.84	2.41	1.70	0.70	16.07	0.02	44.90	0	0.60	0.01

Table 1 - Continued

Fatty Acid Analysis of Leaves from Ma29 Transgenic *Brassica* Plants

Treatment	SPI.	20:02	20:03	20:04	20:05	22:00	22:01	22:02	22:03	22:06	24:0	24:1
	#	%	%	%	%	%	%	%	%	%	%	%
Water	33	0	0	0.29	0	0.01	0.09	16.26	0	0	0.38	0.18
	34	0.01	0	0.26	0	0.14	0.10	16.82	0.02	0.05	0.36	0.27
	35	0.01	0	0.25	0	0.12	0.06	11.29	0.04	0.05	0.29	0.25
	36	0	0.01	0.26	0	0.07	0.04	11.82	0.03	0.36	0.28	0.21
	37	0.02	0	0.21	0	0.18	0.08	15.87	0.06	0.20	0.30	0.17
	38	0.01	0	0.24	0	0.15	0.07	13.64	0.09	0.08	5.89	0.23
GLA	39	0.02	0.01	0.27	0	0.10	0.08	16.25	3.42	0.19	0.37	0.17
	40	0.01	0	0.27	0	0.10	0.10	14.74	0.05	0.10	0.36	0.14
	41	0	0	0.27	0	0.20	0.10	13.15	0.13	0.29	0.33	0.20
	42	0	0	0.28	0	0.11	0.11	12.60	0.02	0.24	0.38	0.13
	43	0.01	0	0.28	0	0.10	0.03	14.73	0.01	0.24	0.34	0.14
	44	0.02	0	0.26	0	0.13	0.07	14.43	0.05	0.16	0.33	0.17
DGLA	45	0.06	1.21	0.26	0	0.07	0.07	18.67	0.02	0.21	0.36	0.13
	46	0	1.94	0.27	0	0.11	0.09	17.97	0.09	0.39	0.41	0.11
	47	0.01	0.69	0.96	0	0.11	0.07	17.96	0	0.22	0.49	0.20
	48	0.01	0.70	0.74	0	0.14	0.09	17.14	0.05	0.32	0.52	0.10
	49	0	0.35	1.11	0	0.10	0.07	17.26	0.07	0.23	0.39	0.18
	50	0	0.20	0.87	0	0.21	0.07	15.73	0.04	0.15	0.37	0.18

Leaves treated with GLA contained from 1.56 to 2.4 wt% GLA. The fatty acid analysis showed that the lipid composition of control and transgenic leaves was essentially the same. Leaves of control plants treated with DGLA contained 1.2-1.9 w% DGLA and background amounts of ARA (.26-.27 wt%). Transgenic leaves contained only .2-.7 wt% DGLA, but levels of ARA were increased (.74-1.1 wt%) indicating that the DGLA was converted to ARA in these leaves.

Expression in Seed

The purpose of this experiment was to determine whether a construct with the seed specific napin promoter would enable expression in seed.

The Ma29 cDNA was modified by PCR to introduce *Xho*I cloning sites upstream and downstream of the start and stop codons, respectively, using the following primers:

Madxho-forward:

5'-CUACUACUACUACTCGAGCAAGATGGGAACGGACCAAGG

Madxho-reverse:

5'-CAUCAUCAUCAUCTCGAGCTACTCTTCCTTGGGACGGAG

The PCR product was subcloned into pAMP1 (GIBCOBRL) using the CloneAmp system (GIBCOBRL) to create pCGN5522 and the Δ5 desaturase sequence was verified by sequencing of both strands.

For seed-specific expression, the Ma29 coding region was cut out of pCGN5522 as an *Xho*I fragment and inserted into the *Sal*I site of the napin expression cassette, pCGN3223, to create pCGN5528. The *Hin*dIII fragment of pCGN5528 containing the napin 5' regulatory region, the Ma29 coding region, and the napin 3' regulatory region was inserted into the *Hin*dIII site of pCGN1557 to create pCGN5531. Two copies of the napin transcriptional unit were inserted in tandem. This tandem construct can permit higher expression of the desaturases per genetic loci. pCGN5531 was introduced into *Brassica napus* cv.LP004 via Agrobacterium mediated transformation.

The fatty acid composition of twenty-seed pools of mature T2 seeds was analyzed by GC. Table 2 shows the results obtained with independent transformed lin s as compared to non-transform d LP004 seed. The transgenic seeds containing

pCGN5531 contain two fatty acids that are not present in the control seeds, tentatively identified as taxoleic acid (5,9-18:2) and pinolenic acid (5,9,12-18:3), based on their elution relative to oleic and linoleic acid. These would be the expected products of Δ5 desaturation of oleic and linoleic acids. No other differences in fatty acid composition were observed in the transgenic seeds.

Example 4

Production of D5-desaturated Fatty Acids in Transgenic Plants

The construction of pCGN5531 (D5-desaturase) and fatty acid composition of T2 seed pools is described in Example 3. This example takes the seeds through one more generation and discusses ways to maximize the D5-desaturated fatty acids.

Example 3 describes the fatty acid composition of T2 seed pools of pCGN5531-transformed *B. napus* cv. LP004 plants. To investigate the segregation of D5-desaturated fatty acids in the T2 seeds and to identify individual plants to be taken on to subsequent generations, half-seed analysis was done. Seeds were germinated overnight in the dark at 30 degrees on water-soaked filter paper. The outer cotyledon was excised for GC analysis and the rest of the seedling was planted in soil. Results of some of these analyses are shown in the accompanying Table 3. D5,9-18:2 accumulated to as high as 12% of the total fatty acids and D5,9,12-18:3 accumulated to up to 0.77% of the fatty acids. These and other individually selected T2 plants were grown in the greenhouse to produce T3 seed.

Table 2
Composition of T2 Pooled Seed

	16:0 %	16:1 %	18:0 %	18:1 %	(5,9)18:2 %	18:2 %	(5,9,12)18:3 %	18:3 %	20:0 %	20:1 %	20:2 %	22:0 %	22:1 %	24:0 %
LP004 control	3.86	0.15	3.05	69.1	0	18.51	0.01	1.65	1.09	1.40	0.03	0.63	0.05	0.42
5531-1	4.26	0.15	3.23	62.33	4.07	21.44	0.33	1.38	0.91	1.04	0.05	0.41	0.03	0.27
5531 2	3.78	0.14	3.37	66.18	4.57	17.31	0.27	1.30	1.03	1.18	0	0.47	0.01	0.30
5531-6	3.78	0.13	3.47	63.61	6.21	17.97	0.38	1.34	1.04	1.14	0.05	0.49	0.02	0.26
5531-10	3.96	0.17	3.28	63.82	5.41	18.58	0.32	1.43	0.98	1.11	0.02	0.50	0	0.31
5531-16	3.91	0.17	3.33	64.31	5.03	18.98	0.33	1.39	0.96	1.11	0	0.44	0	0
5531-28	3.81	0.13	2.58	62.64	5.36	20.95	0.45	1.39	0.83	1.15	0.01	0.36	0.05	0.21

Table 3

Fatty acid analysis of selected T2 half-seeds from pCGN5531-LP004 events

CYCLE ID	SPL NO	STRAIN ID	12:0	14:0	16:0	16:1	18:0	18:1	18:2 Δ5,9	18:2 Δ9,12	18:3 Δ5,9,12	18:3 Δ9,12,15
97XX1539	93	5531-LP004-6	0.03	0.07	3.92	0.17	3.5	61.32	12.22	15.36	0.77	1.36
97XX1539	29	5531-LP004-6	0.01	0.04	3.6	0.09	3.23	63.77	10.63	14.47	0	1.22
97XX1539	38	5531-LP004-6	0.01	0.05	3.71	0.09	3.02	65.13	10.57	13.98	0	1.06
97XX1539	41	5531-LP004-6	0.01	0.05	3.64	0.07	3.22	62.51	9.7	16.63	0	1.28
97XX1539	18	5531-LP004-6	0.02	0.06	3.69	0.09	3.36	63.79	9.63	15.29	0.63	1.15
97XX1539	85	5531-LP004-6	0.01	0.06	3.6	0.09	3.54	64.81	9.54	13.69	0.6	1.26
98GC0023	98	5531-LPUU4-23	0.01	0.05	3.5	0.09	3.12	64.97	9.92	13.62	0.55	1.25
98GC0023	32	5531-LP004-23	0.01	0.05	3.43	0.08	2.62	65.21	9.83	14.28	0.59	1.15
98GC0023	78	5531-LP004-23	0.01	0.05	3.45	0.07	2.78	64.97	9.34	14.69	0.58	1.17
98GC0023	86	5531-LP004-23	0.01	0.05	3.32	0.08	2.7	64.18	9.08	15.99	0.68	1.18
98GC0023	67	5531-LP004-23	0.01	0.04	3.49	0.08	3.03	64.14	8.78	15.95	0.62	1.08
98GC0023	52	5531-LP004-23	0.01	0.03	3.38	0.07	2.56	67.44	8.65	13.55	0.5	1.02

To maximize the accumulation of D5,9 18:2 in seed oil, the pCGN5531 construct could be introduced into a high oleic acid variety of canola. A high-oleic variety could be obtained by mutation, so-suppression, or antisense suppression of the D12 and D15 desaturases or other necessary co-factors.

To maximize accumulation of D5,9,12 18:3 in canola, the pCGN5531 construct could be introduced into a high linoleic strain of canola. This could be achieved by crossing pCGN5531-transformed plants with pCGN5542-(M. alpina D12-desaturase) transformed plants. Alternatively, the D5 and D12 desaturases could be combined on one T-DNA for transformation. The transcriptional unit consisting of the napin 5'-regulatory region, the *M. alpina* D12-desaturase coding region, and the napin 3'-regulatory region can be cut out of pCGN5541 (described in CGAB320) as a NotI fragment. NotI/XbaI linkers could be ligated and the resulting fragment insdrted into the XbaI site of pCGN5531. The resulting plasmid would contain three napin transcriptional units containing the *M. alpina* D12 desaturase, and two copies of the napin/*M. alphina* D5 desaturase/napin unit, all oriented in the same direction as the 35S/nptII/tml transcriptional unit used for selection of transformed tissue.

Example 5

Expression of *M. alpina* Δ6 Desaturase in *Brassica napus*

A nucleic acid sequence from a partial cDNA clone, Ma524, encoding a Δ6 fatty acid desaturase from *Mortierella alpina* was obtained by random sequencing of clones from the *M. alpina* cDNA library. The Ma524 cDNA was modified by PCR to introduce cloning sites using the following primers:

Ma524PCR-1

5'-CUACUACUACUATCTAGACTCGAGACCATGGCTGCTGCT
CCAGTGTG

Ma524PCR-2

5'-CAUCAUCAUCAUAGGCCTCGAGTTACTGCGCCTTACCCAT

These primers allowed the amplification of the entire coding region and added *XbaI* and *XhoI* sites to the 5'-end and *XhoI* and *StuI* sites to the 3' end. The PCR product was subcloned into pAMP1 (GIBCOBRL) using the CloneAmp system (GIBCOBRL) to create pCGN5535 and the Δ6 desaturase sequence was verified by sequencing of both strands.

For seed-specific expression, the Ma524 coding region was cut out of pCGN5535 as an *XhoI* fragment and inserted into the *SalI* site of the napin expression cassette, pCGN3223, to create pCGN5536. The *NotI* fragment of pCGN5536 containing the napin 5' regulatory region, the Ma524 coding region, and the napin 3' regulatory region was inserted into the *NotI* site of pCGN1557 to create pCGN5538. pCGN5538 was introduced into *Brassica napus* cv.LP004 via *Agrobacterium* mediated transformation.

Maturing T2 seeds were collected from 6 independent transformation events in the greenhouse. The fatty acid composition of single seeds was analyzed by GC. Table 4 shows the results of control LP004 seeds and six 5538 lines. All of the 5538 lines except #8 produced seeds containing GLA. Presence of GLA segregated in these seeds as is expected for the T2 selfed seed population. In addition to GLA, the *M. alpina* Δ6 desaturase is capable of producing 18:4 (stearidonic) and another fatty acid believed to be the 6,9-18:2.

The above results show that desaturases with three different substrate specificities can be expressed in a heterologous system and used to produce poly-unsaturated long chain fatty acids. Exemplified were the production of ARA (20:4) from the precursor 20:3 (DGLA), the production of GLA (18:3) from 18:2 substrate, and the conversion of 18:1 substrate to 18:2, which is the precursor for GLA.

Table 4

Fatty Acid Analysis of Seeds from Ma524 Transgenic *Brassica* Plants

SPL		16:0	16:1	18:0	18:1	6,9 18:2	18:2	18:3ga	18:3	18:4	20:1	22:0	22:1	24:0	24:1
#		%	%	%		%	%	%	%	%	%	%	%	%	%
	LPOO4-1	4.33	0.21	3.78	72.49	0	13.97	0	1.7	0	1.34	0.71	0.02	0.58	0.27
	-2	4.01	0.16	3.09	73.59	0	14.36	0.01	1.4	0	1.43	0.66	0.02	0.5	0.2
	-3	4.12	0.19	3.56	70.25	0	17.28	0	1.57	0	1.28	0.5	0.02	0.39	0.2
	-4	4.22	0.2	2.7	70.25	0	17.86	0	1.61	0	1.31	0.53	0.02	0.4	0.24
	-5	4.02	0.16	3.41	72.91	0	14.45	0.01	1.45	0	1.37	0.7	0.02	0.51	0.26
	-6	4.22	0.18	3.23	71.47	0	15.92	0.01	1.52	0	1.32	0.69	0.02	0.51	0.27
	-7	4.1	0.16	3.47	72.06	0	15.23	0	1.52	0	1.32	0.63	0.03	0.49	0.23
	-9	4.01	0.17	3.71	72.98	0	13.97	0.01	1.41	0	1.45	0.74	0.03	0.58	0.23
	-10	4.04	0.16	3.57	70.03	0	17.46	0	1.5	0	1.33	0.61	0.03	0.36	0.24
	5538-1-1	4.61	0.2	3.48	68.12	1.37	10.68	7.48	1.04	0.33	1.19	0.49	0.02	0.33	0.13
	-2	4.61	0.22	3.46	68.84	1.36	10.28	7.04	1.01	0.31	1.15	0.48	0.02	0.39	0
	-3	4.76	0.24	3.24	65.86	0	21.36	0	1.49	0	1.08	0.46	0.02	0.38	0.22
	-4	4.84	0.3	3.89	67.64	1.67	9.9	6.97	1.02	0.36	1.14	0.53	0.02	0.5	0.18
	-5	4.64	0.2	3.58	64.5	3.61	8.85	10.14	0.95	0.48	1.19	0.47	0.01	0.33	0.12
	-6	4.91	0.27	3.44	66.51	1.48	11.14	7.74	1.15	0.33	1.08	0.49	0.02	0.34	0.13
	-7	4.67	0.22	3.24	65.78	1.27	11.92	8.38	1.2	0	1.12	0.47	0.02	0.37	0.16
	-8	4.59	0.22	3.4	70.77	0	16.71	0	1.35	0	1.14	0.48	0.02	0.39	0.15
	-9	4.63	0.23	3.51	69.66	2.01	8.77	7.24	0.97	0	1.18	0.52	0.02	0.3	0.11
	-10	4.56	0.19	3.55	70.68	0	16.89	0	1.37	0	1.22	0.54	0.02	0.22	0.03
	5538-3-1	4.74	0.21	3.43	67.52	1.29	10.91	7.77	1.03	0.28	1.11	0.5	0.02	0.35	0.14
	-2	4.72	0.21	3.24	67.42	1.63	10.37	8.4	0.99	0	1.12	0.49	0.02	0.36	0.15
	-3	4.24	0.21	3.52	71.31	0	16.53	0	1.33	0	1.12	0.45	0.02	0.4	0.14

Table 4

Fatty Acid Analysis of Seeds from Ma524 Transgenic *Brassica* Plants

SPL		16:0	16:1	18:0	18:1	6.9 18:2	18:2	18:3ga	18:3	18:4	20:1	22:0	22:1	24:0	24:1
#		%	%	%		%	%	%	%	%	%	%	%	%	%
	-4	4.64	0.21	3.45	67.92	1.65	9.91	7.97	0.91	0.33	1.14	0.47	0.02	0.37	0.14
	-5	4.91	0.25	3.31	67.19	0	19.92	0.01	1.39	0	1.05	0.48	0.02	0.37	0.14
	-6	4.67	0.21	3.25	67.07	1.23	11.32	8.35	0.99	0	1.16	0.47	0.02	0.33	0.16
	-7	4.53	0.19	2.94	64.8	4.94	8.45	9.95	0.93	0.44	1.13	0.37	0.01	0.27	0.12
	-8	4.66	0.22	3.68	67.33	0.71	12	6.99	1.1	0.24	1.18	0.48	0.03	0.36	0.17
	-9	4.65	0.24	3.11	67.42	0.64	12.71	6.93	1.16	0.25	1.08	0.45	0.02	0.32	0.17
	-10	4.88	0.27	3.33	65.75	0.86	12.89	7.7	1.1	0.24	1.08	0.46	0.01	0.34	0.16
	5538-4-1	4.65	0.24	3.8	62.41	0	24.68	0	1.6	0.01	0.99	0.45	0.02	0.33	0.13
	-2	5.37	0.31	3	57.98	0.38	18.04	10.5	1.41	0	0.99	0.48	0.02	0.3	0.19
	-3	4.61	0.22	3.07	63.62	0.3	16.46	7.67	1.2	0	1.18	0.45	0.02	0.29	0.14
	-4	4.39	0.19	2.93	65.97	0	22.36	0	1.45	0	1.17	0.41	0.03	0.32	0.15
	-5	5.22	0.29	3.85	62.1	2.35	10.25	11.39	0.93	0.41	1.04	0.6	0.02	0.47	0.17
	-6	4.66	0.18	2.85	66.79	0.5	13.03	7.66	0.97	0.22	1.28	0.42	0.02	0.31	0.14
	-7	4.85	0.26	3.03	57.43	0.26	28.04	0.01	2.59	0.01	1.13	0.56	0.02	0.4	0.23
	-8	5.43	0.28	2.94	54.8	1.84	13.79	15.67	1.36	0.53	1.1	0.55	0.02	0.35	0.19
	-9	4.88	0.24	3.32	62.3	0.58	14.86	9.04	1.34	0.29	1.13	0.52	0.02	0.37	0.19
	-10	4.53	0.2	2.73	64.2	0.07	24.15	0	1.52	0	1.09	0.39	0.02	0.27	0.17
	5538-5-1	4.5	0.15	3.35	66.71	0.88	11.7	8.38	1.04	0.3	1.24	0.49	0.02	0.29	0.17
	-2	4.77	0.23	3.06	62.67	0.68	15.2	8.8	1.31	0.28	1.15	0.46	0.02	0.3	0.19
	-3	4.59	0.72	3.61	64.35	2.29	9.95	10.57	1.01	0.45	1.21	0.48	0.02	0.26	0.16
	-4	4.86	0.26	3.4	67.69	0.65	12.24	6.61	1.09	0.23	1.07	0.45	0.02	0.32	0.14
	-5	4.49	0.21	3.3	69.25	0.04	16.51	2.18	1.2	0	1.11	0.44	0.02	0.33	0.16

Table 4

Fatty Acid Analysis of Seeds from Ma524 Transgenic *Brassica* Plants

SPL		16:0	16:1	18:0	18:1	6,9 18:2	18:2	18:3ga	18:3	18:4	20:1	22:0	22:1	24:0	24:1
#		%	%	%		%	%	%	%	%	%	%	%	%	%
	-6	4.5	0.21	3.47	70.48	0.08	14.9	2.19	1.22	0	1.13	0.49	0.02	0.33	0.16
	-7	4.39	0.21	3.44	67.59	2.38	9.24	8.98	0.89	0	1.18	0.44	0.02	0.28	0.14
	-8	4.52	0.22	3.17	68.33	0.01	18.91	0.73	1.32	0.01	1.08	0.45	0.02	0.29	0.17
	-9	4.68	0.2	3.05	64.03	1.93	11.03	11.41	1.02	0.01	1.15	0.39	0.02	0.21	0.15
	-10	4.57	0.2	3.1	67.21	0.61	12.62	7.68	1.07	0.25	1.14	0.43	0.02	0.25	0.15
	5538-8-1	4.95	0.26	3.14	64.04	0	23.38	0	1.54	0	0.99	0.42	0.02	0.38	0.17
	-2	4.91	0.26	3.71	62.33	0	23.97	0	1.77	0	0.95	0.53	0.02	0.42	0.19
	-3	4.73	0.25	4.04	63.83	0	22.36	0.01	1.73	0	1.05	0.55	0.02	0.45	0.16
	-4	5.1	0.35	3.8	60.45	0	24.45	0.01	2.13	0	1.07	0.65	0.03	0.53	0.24
	-5	4.98	0.3	3.91	62.48	0	23.44	0	1.77	0	1.01	0.51	0.01	0.43	0.21
	-6	4.62	0.21	3.99	66.14	0	20.38	0	1.48	0	1.15	0.53	0.02	0.48	0.19
	-7	4.64	0.22	3.55	64.6	0	22.65	0	1.38	0	1.09	0.45	0.02	0.41	0.19
	8	5.65	0.38	3.18	56.6	0	30.83	0.02	0.02	0	0.98	0.55	0.03	0.39	0.26
	-9	8.53	0.63	6.9	51.76	0	26.01	0	0.01	0	1.41	1.21	0.07	0.96	0.33
	-10	5.52	0.4	3.97	57.92	0	28.95	0	0.02	0	0.95	0.52	0.02	0.41	0.16
	5538 10 1	4.44	0.19	3.5	68.42	0	19.51	0	1.32	0	1.14	0.45	0.02	0.31	0.16
	-2	4.57	0.21	3.07	66.08	0	21.99	0.01	1.36	0	1.12	0.41	0.02	0.31	0.16
	-3	4.63	0.21	3.48	67.43	0	20.27	0.01	1.32	0	1.12	0.46	0.02	0.21	0.08
	-4	4.69	0.19	3.22	64.62	0	23.16	0	1.35	0	1.08	0.46	0.02	0.33	0.2
	-5	4.58	0.2	3.4	68.75	0	20.17	0.01	0.02	0	1.1	0.45	0.02	0.34	0.17
	-8	4.55	0.21	0	73.55	0.05	14.91	2.76	1.21	0.07	1.24	0.51	0.02	0.19	0
	-9	4.58	0.21	3.28	66.19	0	21.55	0	1.35	0	1.12	0.43	0.02	0.33	0.16

Table 4

Fatty Acid Analysis of Seeds from Ma524 Transgenic *Brassica* Plants

SPL		16:0	16:1	18:0	18:1	6,9 18:2	18:2	18:3ga	18:3	18:4	20:1	22:0	22:1	24:0	24:1
#		%	%	%		%	%	%	%	%	%	%	%	%	%
	-10	4.52	0.2	3.4	68.37	0	19.33	0.01	1.3	0	1.13	0.46	0.02	0.35	0.18

Example 6

Production of D6,9 18:2 in Canola Oil

Example 5 described construction of pCGN5538 designed to express the ***M. alpina*** D6 desaturase in seeds of transgenic canola. Table 4 in that example showed examples of single seed analyses from 6 independent transgenic events. Significant amounts of GLA were produced, in addition to the D6,9 18:2 fatty acid.

A total of 29 independent pCGN5538-transformed transgenic plants of the low-linolenic LP004 cultivar were regenerated and grown in the greenhouse. Table 5 shows the fatty acid composition of 20-seed pools of T2 seed from each event. Seven of the lines contained more than 2% of the D6,9 18:2 in the seed pools. To identify and select plants with high amounts of D6,9 18:2 to be taken on to subsequent generations, half-seed analysis was done. Seeds were germinated overnight in the dark at 30 degrees on water-soaked filter paper. The outer cotyledon was excised for GC analysis and the rest of the seedling was planted in soil. Based on results of fatty acid analysis, selected T2 plants were grown in the greenhouse to produce T3 seed. The selection cycle was repeated; pools of T3 seed were analyzed for D6,9 18:2, T3 half-seeds were dissected and analyzed, and selected T3 plants were grown in the greenhouse to produce T4 seed. Pools of T4 seed were analyzed for fatty acid composition. Table 5 summarizes the results of this process for lines derived from one of the original transgenic events, 5538-LP004-25. Levels of D6,9 18:2 have thus been maintained through 3 generations.

To maximize the amount of D6,9 18:2 that could be produced, the pCGN5538 construct could be introduced into a high oleic acid variety of canola either by transformation or crossing. A high-oleic variety could be obtained by mutation, co-suppression, or antisense suppression of the D12 and D15 desaturases or other necessary co-factors.

Table 5

Fatty Acid Composition of 20-seed Pools of pCGN5538 T2 Seeds

								C9	6,9		C6912	C91215						
SPL	5538-LP004	12:0	14:0	16:0	16:1	18:0	18:1	18:1	18:2	18:2	18:3	18:3	18:4	20:0	20:1	20:2	22:0	22:2
#	event	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
	31	0.02	0.06	4.07	0.07	0	59.4		5.4	10.07	15.93	1.2	0.6	0.98	1.16	0.0	0.44	0.03
	29	0.01	0.05	3.81	0.14	0	60.7		4.53	10.9	14.77	1.03	0.55	1.09	1.26	0.0	0.46	0.02
	19	0.02	0.07	4.27	0.13	0	62.9		4.17	10.03	13.14	1.02	0.59	1.18	1.25	0.0	0.53	0.02
	14	0	0	3.29	0.24	3.8	49.1		1.02	23.44	11.21	2.26	0.34	1.45	0.93	0.0	0.76	0
	22	0.02	0.05	3.87	0.09	0	64.1		2.59	12.57	11.18	1.27	0.6	1.18	1.03	0.1	0.56	0
	9	0.01	0.06	4.57	0.16	0	62.9		3.4	12.05	11.15	1.27	0.6	1.28	1.18	0.0	0.56	0.03
	25	0.01	0.06	4.17	0.14	0	62.4		2.49	14.42	11.03	1.2	0.46	1.18	1.15	0.0	0.53	0.01
	15	0.01	0.05	3.94	0.11	0	65.2		2.08	12.77	10.9	1.04	0.43	1.1	1.24	0.0	0.48	0.01
	18	0	0.06	5.34	0.29	0	58.4		1.42	18.19	10.53	1.8	0.49	1.2	1	0.0	0.58	0.02
	20	0.01	0.04	3.95	0.1	0	65.6		1.31	13.83	10.22	1.09	0.39	1.06	1.3	0.0	0.46	0.08
	7	0.02	0.07	4.04	0.11	0	62.1		0.92	18.12	8.72	1.77	0.35	1.26	1.19	0.0	0.58	0
	11	0.01	0.06	4.23	0.17	0	62.9		1.6	17.19	8.58	1.48	0.38	1.16	1.03	0.0	0.49	0.01
	27	0.02	0	3.99	0.14	0	65.3		0.64	17.85	7.89	1.36	0.31	1.08	1.21	0.0	0	0
	2	0.01	0.05	4.02	0.14	0	66.4		1.2	15.74	7.58	1.22	0.32	1.06	1.19	0.0	0.45	0
	28	0.01	0.04	3.77	0.11	0	67.5		0.79	15.56	7.58	1.12	0.28	0.97	1.23	0.0	0.44	0
	3	0.01	0.05	3.96	0.12	0	68.5		1.81	13.23	7.44	1.1	0.35	1.12	1.21	0.0	0.46	0.01
	21	0.01	0.05	3.74	0.1	0	66.9		1.16	15.9	6.97	1.35	0.28	1.15	1.27	0.0	0.52	0
	5	0.01	0.04	3.81	0.12	0	69.1		0.74	14.58	6.95	1.14	0.28	1.06	1.18	0.0	0.45	0
	6	0	0	2.84	0	3.06	62.5		1.55	18.44	6.94	1.21	0.39	1.04	1.33	0.1	0	0
	4	0.01	0.05	3.88	0.11	0	66.9		0.64	16.21	6.89	1.52	0.31	1.09	1.21	0.0	0.5	0
	30	0.01	0.04	3.89	0.12	0	68.6		0.72	15.58	6.47	1.17	0.23	1.03	1.07	0.0	0.46	0
	16	0.02	0.05	3.75	0.13	0	70.4		0.91	13.56	6.39	1.13	0.28	1.04	1.2	0.0	0.44	0.01
	26	0.01	0	3.77	0.12	0	67.6		0	21.08	3.61	1.37	0.13	0.96	1.16	0.0	0	0

PATENT

Table 5

Fatty Acid Composition of 20-seed Pools of PCGN5538 T2 Seeds

							C9	6.9		C6912	C91215						
23	0	0	4.92	0.22	0	65.2		0	22.23	3	1.79	0.11	1.28	1.11	0.0	0	0
24	0.01	0	3.84	0.13	0	68.4		0.36	21	2.09	1.53	0.08	1.06	1.27	0.0	0	0
10	0.01	0	3.74	0.11	0	70.4		0	20.82	0.65	1.3	0.03	1.05	1.18	0.0	0.46	0
12	0.01	0	3.83	0.12	0	69.9		0	21.61	0.34	1.34	0	1.06	1.12	0.0	0.47	0
17	0.01	0	0	0.13	0	72.6		0	23.03	0.34	1.51	0	1.18	1.21	0.0	0	0
8	0.01	0	4.54	0.2	0	64.9		0	25.65	0.22	1.94	0	1.38	1.01	0.0	0	0
13	0.01	0	3.99	0.16	0	65.6		0	25.97	0	1.58	0	1.17	1.16	0.0	0	0
LP004 control	0.01	0.04	3.46	0.09	0	69.9		0	21.93	0	1.37	0.01	0.9	1.25	0.0	0.42	0

Example 7

Identification of potentially useful D15/ω-3 desaturases from other organisms

To look for desaturases involved in PUFA production, cDNA libraries were constructed from total RNA isolated from *Schizochytrium* (unknown species - proprietary strain supplied by Kelco in San Diego). Plasmid-based cDNA libraries were constructed in pSPORT1 (GIBCO-BRL) following manufacturer's instructions using a commercially available kit (GIBCO-BRL). Random cDNA clones were sequenced and nucleic acid sequences that encode putative desaturases were identified through BLAST search of the databases and comparison to known D12 and D15 sequences.

One clone was identified from the *Schizochytrium* library with homology to both D12 and D15 desaturases; it is called 81-53-A2. The DNA Sequence is presented as Seq ID NO:1. The corresponding peptide sequence is presented as SEQ ID NO: 2

SEQUENCE LISTING

(1) GENERAL INFORMATION:

APPLICANT: KNUTZON, DEBORAH et al.

(ii) TITLE OF INVENTION: POLY-UNSATURATED FATTY ACIDS IN PLANTS

(iii) NUMBER OF SEQUENCES:

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: LIMBACH & LIMBACH L.L.P.
(B) STREET: 2001 FERRY BUILDING
(C) CITY: SAN FRANCISCO
(D) STATE: CA
(E) COUNTRY: USA]
(F) ZIP: 94111

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: Microsoft Word

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/833,610
(B) FILING DATE: 11-APR-1997

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: MICHAEL R. WARD

(B) REGISTRATION NUMBER: 38.351

(C) REFERENCE/DOCKET NUMBER: CGMO-100

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (415) 433-4150

(B) TELEFAX: (415) 433-8716

(C) TELEX: N/A

(2) INFORMATION FOR SEQ ID NO:1:

```
CACGGAAGCA AGCCTTGACA TCCTTTGCCA ACATGTGCAA GGTCGAGACC AAGCAGCACG     60
CTGCGGCCAC CGCCGTCCAG GCACCGGAGC AGCAGCAGCA GCAGCAGCAG CAGTCGCAGC    120
AGTCGCACCA GCTGCCGTCC GCAGCGCAAG CCTCGGCGGG GGAAGTCCTG GAAAAACGAC    180
CCATTATCCA TGGCAAGCAC AACCCAGACC TGCCCACGCT CGGAGAGATT CGCGCCGCGG    240
TGCCCAAGCA CTGCTTCCAC CGCTCGCTCC TCACCAGTTC GCTGTATCTT CGGCGCGACC    300
TCGTCATGGC AACGATCCTG TTCTGCATGG CGCGGCACTT CCTGCCCGTA TACGACATGG    360
GCGTCTTGGG CGCCATCGGC TGGACAAGCC TACGTAATTG TTCAAGGGAC AGTCTTTGCT    420
CGACTCTGGG TCCTCGGTCA TGAGTGCGGA CACCAGGCAT TTTCCAACTA CAGAGTAGTC    480
AACGACACCG TCGGCTACCT TGTGCACACT GCCTTGCTTG TGCCTTACTT TAGCTGGGCG    540
TACACGCATG GCTTGCACCA CGCCCGTGTC AACCACATGC TCGACGGCGA GTCTCACACT    600
CCCAACCTGC AAAAGAAAGT CATGGCTAAC TTTCAAAAGT TAGCCGACCT CATGGGCGAC    660
GAGGCCTTTG CTGTCCTCCA CGTCTTTGTT TATCTCCTCC TTCCCTGGCC GCTGTATATC    720
ATCAATGGCA GCGGGGCATC CAAGCGCAAC CACGAGGGTA AGCGGTGGTC AAAGGAGATG    780
CTCAAGCGGC CTAACCACTT CTTGCCCACC TCGGAGCTCT TCCCGGACAA GATGCGTCTC    840
AGTGTGGCAG GCTCTACGGC CGGTGTGCTC GTCGTCATTG CTAGCTTGTG TTACTGGGGT    900
TCCATCGAGG GTTCGCGGAC CGTGCTGCTT CAGTACTTTC TCCCTTACCT TGTTGTGAAC    960
GCCTACCTCA TTGGTTTCAC ATGGATGCAG CATACTCACC AAGATGTCCC GCATCTTGGC   1020
GAAGACGAAG TGGTCCTGGC TCGCTGGAAC CATTCTCACC ATCGATCGCC TTATCCTGCC   1080
TTTATCGACG TACTCACACA CCGCATCGGA TCCACGCACG TTGCGCATCA TCTTTTCTCG   1140
AAGATGCCCT GGTACCATGC CCGCGAAGCC ACTCTTCACA TCAAGGCTTT GCTTGAGCCC   1200
AAGGGGGTCT ACAACTATGA CCCGACGCCG TTTACAAGGC GCTGTACAAC ACCGCCAGAT   1260
ACTGTCACTT TATGGAGCGC GTCGAAGGCA TTCAGTTCTT CAAACACGTT CACGCTCAAT   1320
CTACAAAAGC AAAGGTCCTC TAAGATTTCC TCCTTCTAGG ATAATCACTT TCATTGCTAC   1340
CATACAATAT AACTTCATCG CCCCTTCCCG TAATCAATTT GTCTCTCTTT TC           1392
```

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:

(B) TYPE: aminoacid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(2) INFORMATION FOR SEQ ID NO:2:

MASTTQTCPRSERFAPRCPSTASSARSSPVRCILGATSSWQRSCSAWRGTSCPYTTWASWAPSAGQAYVIVQGTV
FAGLWVLGHECGHQAPSNYRVVNDTVGYLVHTALLVPYFSWAYTHGLHHARVNHMLDGESHTPNLQKKVHANFQK
LADLMGDEAPAVLHVFVYLLLAWPLYIINGSGASKRNHEGKRWSKEMLKRPNHFLPTSELPPDKMRLSVASSTAG
VLVVIASLCYWGSIEGSRTVLLQYPLPYLVVNAYLIGPTWHQHTHQDVPHLGEDEVVLGRWDHSHHRSPYPAFID
VLTHRIGSTHVAHHLFSKMPWYHAREATVHIKALLEPKGVYNYDPTPFTRECTTPPDTVTLWRASKAPSSSNTLT
LNLQKQRSSKISSF

CLAIMS:

1. A method of producing a polyunsaturated fatty acid in a host cell comprising the steps of:
   (A) transforming a host cell with a nucleotide sequence comprising: 1) an expression cassette comprising a transcriptional and translational initiation regulatory region, said expression cassette being joined in reading frame 5' to 2) a DNA sequence encoding a desaturase polypeptide which modulates the production of polyunsaturated fatty acids; and
   (B) culturing said transformed host cell under time and conditions sufficient for the expression of said desaturase polypeptide in said host cell, expression of said desaturase polypeptide resulting in production of polyunsaturated fatty acids by said host cell.

[redacted]
29

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/82, 15/54

A2

(11) International Publication Number: WO 98/46776

(43) International Publication Date: 22 October 1998 (22.10.98)

(21) International Application Number: PCT/US98/07114

(22) International Filing Date: 9 April 1998 (09.04.98)

(30) Priority Data:
60/041,815 11 April 1997 (11.04.97) US

(71) Applicant *(for all designated States except US)*: CALGENE LLC [US/US]; 1920 Fifth Street, Davis, CA 95616 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: DEHESH, Katayoon [US/US]; 521 Crownpointe Circle, Vacaville, CA 95687 (US).

(74) Agent: SCHWEDLER, Carl, J.; Calgene LLC, 1920 Fifth Street, Davis, CA 95616 (US).

(81) Designated States: AU, BR, CA, JP, KR, MX, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: PLANT FATTY ACID SYNTHASES AND USE IN IMPROVED METHODS FOR PRODUCTION OF MEDIUM–CHAIN FATTY ACIDS

(57) Abstract

By this invention, compositions and methods of use related to β–ketoacyl–ACP synthase of special interest are synthases obtainable from *Cuphea* species. Amino acid and nucleic acid for synthase protein factors are provided, as well as methods to utilize such sequences in constructs for production of genetically engineered plants having altered fatty acid compositions. Of particular interest is the expression of synthase protein factors in conjunction with expression of plant medium–chain acyl–ACP thioesterases for production of increased levels and/or modified ratios of medium–chain fatty acids in oils of transgenic plant seeds.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

PLANT FATTY ACID SYNTHASES AND USE IN IMPROVED METHODS FOR PRODUCTION OF MEDIUM-CHAIN FATTY ACIDS

INTRODUCTION

Field of Invention

The present invention is directed to genes encoding plant fatty acid synthase enzymes relevant to fatty acid synthesis in plants, and to methods of using such genes in combination with genes encoding plant medium-chain preferring thioesterase proteins. Such uses provide a method to increase the levels of medium-chain fatty acids that may be produced in seed oils of transgenic plants.

Background

Higher plants synthesize fatty acids via a common metabolic pathway. In developing seeds, where fatty acids attached to triglycerides are stored as a source of energy for further germination, the fatty acid synthesis pathway is located in the plastids. The first step is the formation of acetyl-ACP (acyl carrier protein) from acetyl-CoA and ACP catalyzed by a short chain preferring condensing enzyme, ß-ketoacyl-ACP synthase (KAS) III. Elongation of acetyl-ACP to 16- and 18- carbon fatty acids involves the cyclical action of the following sequence of reactions: condensation with a two-carbon unit from malonyl-ACP to form a longer ß-ketoacyl-ACP (ß-ketoacyl-ACP synthase), reduction of the

keto-function to an alcohol (ß-ketoacyl-ACP reductase), dehydration to form an enoyl-ACP (ß-hydroxyacyl-ACP dehydrase), and finally reduction of the enoyl-ACP to form the elongated saturated acyl-ACP (enoyl-ACP reductase). ß-ketoacyl-ACP synthase I (KAS I), is primarily responsible for elongation up to palmitoyl-ACP (C16:0), whereas ß-ketoacyl-ACP synthase II (KAS II) is predominantly responsible for the final elongation to stearoyl-ACP (C18:0).

Genes encoding peptide components of ß-ketoacyl-ACP synthases I and II have been cloned from a number of higher plant species, including castor (*Ricinus communis*) and *Brassica* species (USPN 5,510,255). KAS I activity was associated with a single synthase protein factor having an approximate molecular weight of 50 kD (synthase factor B) and KAS II activity was associated with a combination of two synthase protein factors, the 50 kD synthase factor B and a 46 kd protein designated synthase factor A. Cloning and sequence of a plant gene encoding a KAS III protein has been reported by Tai and Jaworski (*Plant Physiol.* (1993) *103*:1361-1367).

The end products of plant fatty acid synthetase activities are usually 16- and 18-carbon fatty acids. There are, however, several plant families that store large amounts of 8- to 14-carbon (medium-chain) fatty acids in their oilseeds. Recent studies with *Umbellularia californica* (California bay), a plant that produces seed oil rich in lauric acid (12:0), have demonstrated the existence of a medium-chain-specific isozyme of acyl-ACP thioesterase

in the seed plastids. Subsequent purification of the 12:0-ACP thioesterase from *Umbellularia californica* led to the cloning of a thioesterase cDNA which was expressed in seeds of *Arabidopsis* and *Brassica* resulting in a substantial accumulation of lauric acid in the triglyceride pools of these transgenic seeds (USPN 5,512,482). These results and subsequent studies with medium-chain thioesterases from other plant species have confirmed the chain-length-determining role of acyl-ACP thioesterases during de novo fatty acid biosynthesis (T. Voelker (1996) *Genetic Engineering*, Ed. J. K. Setlow, Vol. 18, pgs. 111-133).

DESCRIPTION OF THE FIGURES

Figure 1. DNA and translated amino acid sequence of *Cuphea hookeriana* KAS factor B clone chKAS B-2 are provided.

Figure 2. DNA and translated amino acid sequence of *Cuphea hookeriana* KAS factor B clone chKAS B-31-7 are provided.

Figure 3. DNA and translated amino acid sequence of *Cuphea hookeriana* KAS factor A clone chKAS A-2-7 are provided.

Figure 4. DNA and translated amino acid sequence of *Cuphea hookeriana* KAS factor A clone chKAS A-1-6 are provided.

Figure 5. DNA and translated amino acid sequence of *Cuphea pullcherrima* KAS factor B clone cpuKAS B/7-8 are provided.

Figure 6. DNA and translated amino acid sequence of *Cuphea pullcherrima* KAS factor B clone cpuKAS B/8-7A are provided.

Figure 7. DNA and translated amino acid sequence of *Cuphea pullcherrima* KAS factor A clone cpuKAS A/p7-6A are provided.

Figure 8. Preliminary DNA sequence of *Cuphea pullcherrima* KAS factor A clone cpuKAS A/p8-9A is provided.

Figure 9. DNA and translated amino acid sequence of *Cuphea hookeriana* KASIII clone chKASIII-27 are provided.

Figure 10. The activity profile for purified cpuKAS B/8-7A using various acyl-ACP substrates is provided.

Figure 11. The activity profile for purified chKAS A-2-7 and chKAS A-1-6 using various acyl-ACP substrates is provided.

Figure 12. The activity profile for purified castor KAS factor A using various acyl-ACP substrates is provided.

Figure 13. The activity profile for purified castor KAS factor B using various acyl-ACP substrates is provided.

Figure 14. A graph showing the number of plants arranged according to C8:0 content for transgenic plants containing CpFatB1 versus transgenic plants containing CpFatB1 + chKAS A-2-7 is provided.

Figure 15. Graphs showing the %C10/%C8 ratios in transgenic plants containing ChFatB2 (4804-22-357) and in plants resulting from crosses between 4804-22-357 and 5401-9 (chKAS A-2-7 plants) are provided.

Figure 16. Graphs showing the %C10 + %C8 contents in transgenic plants containing ChFatB2 (4804-22-357) and in plants resulting from crosses between 4804-22-357 and 5401-9 (chKAS A-2-7 plants) are provided.

Figure 17. Graphs showing the %C10/%C8 ratios in transgenic plants containing ChFatB2 (4804-22-357) and in plants resulting from crosses between 4804-22-357 and 5413-17 (chKAS A-2-7 + CpFatB1 plants) are provided.

Figure 18. Graphs showing the %C10 + %C8 contents in transgenic plants containing ChFatB2 (4804-22-357) and in

plants resulting from crosses between 4804-22-357 and 5413-17 (chKAS A-2-7 + CpFatB1 plants) are provided.

Figure 19. Graphs showing the %C12:0 in transgenic plants containing Uc FatB1 (LA86DH186) and in plants resulting from crosses with wild type (X WT) and with lines expressing Ch KAS A-2-7.

Figure 20. Graph showing the relative proportions of C12:0 and C14:0 fatty acids in the seeds of transgenic plants containing Uc FatB1 (LA86DH186) and in plants resulting from crosses with wild type (X WT) and with lines expressing Ch KAS A-2-7.

Figure 21. Graphs showing the %C18:0 in transgenic plants containing Garm FatB1 (5266) and in seeds of plants resulting from crosses with wild type (X WT) and with lines expressing Ch KAS A-2-7.

Figure 22. The activity profile of Ch KAS A in protein extracts from transgenic plants containing Ch KAS A-2-7. Extracts were preptreated with the indicated concentrations of cerulenin.

SUMMARY OF THE INVENTION

By this invention, compositions and methods of use related to ß-ketoacyl-ACP synthase (KAS) are provided. Also of interest are methods and compositions of amino acid and nucleic acid sequences related to biologically active plant synthase(s).

In particular, genes encoding KAS protein factors A and B from *Cuphea* species are provided. The KAS genes are of interest for use in a variety of applications, and may be

used to provide synthase I and/or synthase II activities in transformed host cells, including bacterial cells, such as *E. coli*, and plant cells. Synthase activities are distinguished by the preferential activity towards longer and shorter acyl-ACPs as well as by the sensitivity towards the KAS specific inhibitor, cerulenin. Synthase protein preparations having preferential activity towards medium chain length acyl-ACPs are synthase I-type or KAS I. The KAS I class is sensitive to inhibition by cerulenin at concentrations as low as 1μM. Synthases having preferential activity towards longer chain length acyl-ACPs are synthase II-type or KAS II. The KAS enzymes of the II-type are also sensitive to cerulenin, but at higher concentrations (50μM). Synthase III-type enzymes have preferential activity towards short chain length acyl-ACPs and are insensitive to cerulenin inhibition.

Nucleic acid sequences encoding a synthase protein may be employed in nucleic acid constructs to modulate the amount of synthase activity present in the host cell, especially the relative amounts of synthase I-type, synthase II-type and synthase III-type activity when the host cell is a plant host cell. In addition, nucleic acid constructs may be designed to decrease expression of endogenous synthase in a plant cell as well. One example is the use of an anti-sense synthase sequence under the control of a promoter capable of expression in at least those plant cells which normally produce the enzyme.

Of particular interest in the present invention is the coordinate expression of a synthase protein with the

expression of thioesterase proteins. For example, coordinated expression of synthase factor A and a medium-chain thioesterase provides a method for increasing the level of medium-chain fatty acids which may be harvested from transgenic plant seeds. Furthermore, coordinated expression of a synthase factor A gene with plant medium-chain thioesterase proteins also provides a method by which the ratios of various medium-chain fatty acids produced in a transgenic plant may be modified. For example, by expression of a synthase factor A, it is possible to increase the ratio of C10/C8 fatty acids which are produced in plant seed oils as the result of expression of a thioesterase having activity on C8 and C10 fatty acids.

DETAILED DESCRIPTION OF THE INVENTION

A plant synthase factor protein of this invention includes a sequence of amino acids or polypeptide which is required for catalyzation of a condensation reaction between an acyl-ACP having a chain length of C_2 to C_{16} and malonyl-ACP in a plant host cell. A particular plant synthase factor protein may be capable of catalyzing a synthase reaction in a plant host cell (for example as a monomer or homodimer) or may be one component of a multiple peptide enzyme which is capable of catalyzing a synthase reaction in a plant host cell, i.e. one peptide of a heterodimer.

Synthase I (KAS I) demonstrates preferential activity towards acyl-ACPs having shorter carbon chains, C_2-C_{14} and is sensitive to inhibition by cerulenin at concentrations of 1μM. Synthase II (KAS II) demonstrates preferential

activity towards acyl-ACPs having longer carbon chains, C_{14}-C_{16}, and is inhibited by concentrations of cerulenin (50µM). Synthase III demonstrates preferential activity towards acyl-CoAs having very short carbon chains, C_2 to C_6, and is insensitive to inhibition by cerulenin.

Synthase factors A and B, and synthase III proteins obtained from medium-chain fatty acid producing plant species of the genus *Cuphea* are described herein. As described in the following Examples, synthase A from *C. hookeriana* is naturally expressed at a high level and only in the seeds. *C. hookeriana* synthase B is expressed at low levels in all tissues examined. Expression of synthase A and synthase B factors in *E. coli* and purification of the resulting proteins is employed to determine activity of the various synthase factors. Results of these analyses indicate that synthase factor A from *Cuphea hookeriana* has the greatest activity on 6:0-ACP substrates, whereas synthase factor B from *Cuphea pullcherrima* has greatest activity on 14:0-ACP. Similar studies with synthase factors A and B from castor demonstrate similar activity profiles between the factor B synthase proteins from *Cuphea* and castor. The synthase A clone from castor, however, demonstrates a preference for 14:0-ACP substrate.

Expression of a *Cuphea hookeriana* KAS A protein in transgenic plant seeds which normally do not produce medium-chain fatty acids does not result in any detectable modification of the fatty acid types and contents produced in such seeds. However, when *Cuphea hookeriana* KAS A protein is expressed in conjunction with expression of a

medium-chain acyl-ACP thioesterase capable of providing for production of C8 and C10 fatty acids in plant seed oils, increases in the levels of medium-chain fatty acids over the levels obtainable by expression of the medium-chain thioesterase alone are observed. In addition, where significant amounts of C8 and C10 fatty acids are produced as the result of medium-chain thioesterase expression, co-expression of a *Cuphea* KAS A protein also results in an alteration of the proportion of the C8 and C10 fatty acids that are obtained. For example, an increased proportion of C10 fatty acids may be obtained by co-expression of *Cuphea hookeriana* ChFatB2 thioesterase and a chKAS A synthase factor proteins.

Furthermore, when *Cuphea hookeriana* KAS A protein is expressed in conjunction with expression of a medium-chain acyl-ACP thioesterase capable of providing for production of C12 fatty acids in plant seed oils, increases in the levels of medium-chain fatty acids over the levels obtainable by expression of the medium-chain thioesterase alone are also observed. In addition, where significant amounts of C12 and C14 fatty acids are produced as the result of medium-chain thioesterase expression, co-expression of a *Cuphea* KAS A protein also results in an alteration of the proportion of the C12 and C14 fatty acids that are obtained. For example, an increased proportion of C12 fatty acids may be obtained by co-expression of *Uc* FatB1 thioesterase and a chKAS A synthase factor proteins.

However, when *Cuphea hookeriana* KAS A protein is expressed in conjunction with the expression of a long-chain

acyl-ACP thioesterase capable of providing for production of C18 and C18:1 fatty acids in plant seed oils, no effect on the production of long chain fatty acids was observed. Furthermore, when plants transformed to express a long chain acyl-ACP thioesterase from mangosteen (*Garm*FatA1, U.S. Patent Application No. 08/440,845), which preferentially hydrolyzes C18:0 and C18:1 fatty acyl-ACPs, are crossed with nontransformed control plants, a significant reduction in the levels of C18:0 is obtained. Similar reductions are also observed in the levels of C18:0 in the seeds of plants resulting from crosses between plants transformed to express the *Garm*FatA1 and plants expressing the *Cuphea hookeriana* KAS A protein.

Thus, the instant invention provides methods of increasing and/or altering the medium-chain fatty acid compositions in transgenic plant seed oils by co-expression of medium-chain acyl-ACP thioesterases with synthase factor proteins. Furthermore, various combinations of synthase factors and medium-chain thioesterases may be achieved depending upon the particular fatty acids desired. For example, for increased production of C14 fatty acids, synthase protein factors may be expressed in combination with a C14 thioesterase, for example from *Cuphea palustris* or nutmeg may be employed (WO 96/23892). In addition, thioesterase expression may be combined with a number of different synthase factor proteins for additional effects on medium-chain fatty acid composition.

Synthases of use in the present invention include modified amino acid sequences, such as sequences which have

been mutated, truncated, increased and the like, as well as such sequences which are partially or wholly artificially synthesized. The synthase protein encoding sequences provided herein may be employed in probes for further screening or used in genetic engineering constructs for transcription or transcription and translation in host cells, especially plant host cells. One skilled in the art will readily recognize that antibody preparations, nucleic acid probes (DNA and RNA) and the like may be prepared and used to screen and recover synthases and/or synthase nucleic acid sequences from other sources. Typically, a homologously related nucleic acid sequence will show at least about 60% homology, and more preferably at least about 70% homology, between the *R. communis* synthase and the given plant synthase of interest, excluding any deletions which may be present. Homology is determined upon comparison of sequence information, nucleic acid or amino acid, or through hybridization reactions.

Recombinant constructs containing a nucleic acid sequence encoding a synthase protein factor or nucleic acid sequences encoding a synthase protein factor and a medium-chain acyl-ACP thioesterase may be prepared by methods well known in the art. Constructs may be designed to produce synthase in either prokaryotic or eukaryotic cells. The increased expression of a synthase in a plant cell, particularly in conjunction with expression of medium-chain thioesterases, or decreasing the amount of endogenous synthase observed in plant cells are of special interest.

Synthase protein factors may be used, alone or in combination, to catalyze the elongating condensation reactions of fatty acid synthesis depending upon the desired result. For example, rate influencing synthase activity may reside in synthase I-type, synthase II-type, synthase III-type or in a combination of these enzymes. Furthermore, synthase activities may rely on a combination of the various synthase factors described herein.

Constructs which contain elements to provide the transcription and translation of a nucleic acid sequence of interest in a host cell are "expression cassettes". Depending upon the host, the regulatory regions will vary, including regions from structural genes from viruses, plasmid or chromosomal genes, or the like. For expression in prokaryotic or eukaryotic microorganisms, particularly unicellular hosts, a wide variety of constitutive or regulatable promoters may be employed. Among transcriptional initiation regions which have been described are regions from bacterial and yeast hosts, such as *E. coli, B. subtilis, Saccharomyces cerevisiae*, including genes such as ß-galactosidase, T7 polymerase, trp-lac (tac), trp E and the like.

An expression cassette for expression of synthase in a plant cell will include, in the 5' to 3' direction of transcription, a transcription and translation initiation control regulatory region (also known as a "promoter") functional in a plant cell, a nucleic acid sequence encoding a synthase, and a transcription termination region. Numerous transcription initiation regions are available

which provide for a wide variety of constitutive or regulatable, e.g., inducible, transcription of the desaturase structural gene. Among transcriptional initiation regions used for plants are such regions associated with cauliflower mosaic viruses (35S, 19S), and structural genes such as for nopaline synthase or mannopine synthase or napin and ACP promoters, etc. The transcription/ translation initiation regions corresponding to such structural genes are found immediately 5' upstream to the respective start codons. Thus, depending upon the intended use, different promoters may be desired.

Of special interest in this invention are the use of promoters which are capable of preferentially expressing the synthase in seed tissue, in particular, at early stages of seed oil formation. Examples of such seed-specific promoters include the region immediately 5' upstream of a napin or seed ACP genes such as described in USPN 5,420,034, desaturase genes such as described in Thompson *et al* (*Proc. Nat. Acad. Sci.* (1991) *88:2578-2582)*, or a Bce-4 gene such as described in USPN 5,530,194. Alternatively, the use of the 5' regulatory region associated with the plant synthase structural gene, i.e., the region immediately 5' upstream to a plant synthase structural gene and/or the transcription termination regions found immediately 3' downstream to the plant synthase structural gene, may often be desired. In general, promoters will be selected based upon their expression profile which may change given the particular application.

In addition, one may choose to provide for the transcription or transcription and translation of one or more other sequences of interest in concert with the expression or anti-sense of the synthase sequence, particularly medium-chain plant thioesterases such as described in USPN 5,512,482, to affect alterations in the amounts and/or composition of plant oils.

When one wishes to provide a plant transformed for the combined effect of more than one nucleic acid sequence of interest, a separate nucleic acid construct may be provided for each or the constructs may both be present on the same plant transformation construct. The constructs may be introduced into the host cells by the same or different methods, including the introduction of such a trait by crossing transgenic plants via traditional plant breeding methods, so long as the resulting product is a plant having both characteristics integrated into its genome.

Normally, included with the DNA construct will be a structural gene having the necessary regulatory regions for expression in a host and providing for selection of transformed cells. The gene may provide for resistance to a cytotoxic agent, e.g. antibiotic, heavy metal, toxin, etc., complementation providing prototrophy to an auxotrophic host, viral immunity or the like. Depending upon the number of different host species into which the expression construct or components thereof are introduced, one or more markers may be employed, where different conditions for selection are used for the different hosts.

The manner in which the DNA construct is introduced into the plant host is not critical to this invention. Any method which provides for efficient transformation may be employed. Various methods for plant cell transformation include the use of Ti- or Ri-plasmids, microinjection, electroporation, liposome fusion, DNA bombardment or the like. In many instances, it will be desirable to have the construct bordered on one or both sides by T-DNA, particularly having the left and right borders, more particularly the right border. This is particularly useful when the construct uses *A. tumefaciens* or *A. rhizogenes* as a mode for transformation, although the T-DNA borders may find use with other modes of transformation.

The expression constructs may be employed with a wide variety of plant life, particularly plant life involved in the production of vegetable oils. These plants include, but are not limited to rapeseed, peanut, sunflower, safflower, cotton, soybean, corn and oilseed palm.

For transformation of plant cells using *Agrobacterium*, explants may be combined and incubated with the transformed *Agrobacterium* for sufficient time for transformation, the bacteria killed, and the plant cells cultured in an appropriate selective medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of vegetable oils.

The invention now being generally described, it will be more readily understood by reference to the following examples which are included for purposes of illustration only and are not intended to limit the present invention.

EXAMPLES

Example 1 Cuphea KAS Factor A and B Gene Cloning

Total RNA isolated from developing seeds of *Cuphea hookeriana* and *Cuphea pullcherrima* was used for cDNA synthesis in commercial l-based cloning vectors. For cloning each type of KAS gene, approximately 400,000-500,000 unamplified recombinant phage were plated and the plaques transferred to nitrocellulose. For KAS factor B cloning from *C. hookeriana*, a mixed probe containing *Brassica napus* KAS factor B and *Ricinus communis* (Castor) KAS factor B radiolabeled cDNA's was used. Similarly, a mixed probe containing *Brassica napus* KAS factor A and *Ricinus communis* KAS factor A cDNA clones was used to obtain *C. hookeriana* KAS factor A genes. For KASIII, a spinach KASIII cDNA clone obtained from Dr. Jan Jaworski was radiolabeled and used as a probe to isolate a KASIII clone from *C. hookeriana*. For KAS B and KAS A cloning from *C. pullcherrima*, *C. hookeriana* KAS B and KAS A genes chKAS B-2 and chKAS A-2-7 (see below) were radiolabeled and used as probes.

DNA sequence and translated amino acid sequence for *Cuphea* KAS clones are provided in Figures 1-9. *Cuphea hookeriana* KAS factor B clones chKAS B-2 and chKAS B-31-7

are provided in Figures 1 and 2. Neither of the clones is full length. *Cuphea hookeriana* KAS Factor A clones chKAS A-2-7 and chKAS A-1-6 are provided in Figures 3 and 4. chKAS A-2-7 contains the entire encoding sequence for the KAS factor protein. Based on comparison with other plant synthase proteins, the transit peptide is believed to be represented in the amino acids encoded by nucleotides 125-466. chKAS A-1-6 is not a full length clone although some transit peptide encoding sequence is present. Nucleotides 1-180 represent transit peptide encoding sequence, and the mature protein encoding sequence is believed to begin at nucleotide 181.

Cuphea pullcherrima KAS factor B clones cpuKAS B/7-8 and cpuKAS B/8-7A are provided in Figures 5 and 6. Both of the clones contain the entire encoding sequences for the KAS factor B proteins. The first 35 amino acids of cpuKAS B/7-8 are believed to represent the transit peptide, with the mature protein encoding sequence beginning at nucleotide 233. The first 39 amino acids of cpuKAS B/8-7A are believed to represent the transit peptide, with the mature protein encoding sequence beginning at nucleotide 209. *Cuphea pullcherrima* KAS factor A clones cpuKAS A/p7-6A and cpuKAS A-p8-9A are provided in Figures 7 and 8. Both of the clones contain the entire encoding sequences for the KAS factor A proteins. Translated amino acid sequence of cpuKAS A/p7-6A is provided. The mature protein is believed to begin at the lysine residue encoded 595-597, and the first 126 amino acids are believed to represent the transit peptide. The DNA sequence of KAS A clone cpuKAS A-p8-9A is preliminary.

Further analysis will be conducted to determine final DNA sequence and reveal the amino acid sequence encoded by this gene.

DNA and translated amino acid sequence of *Cuphea hookeriana* KASIII clone chKASIII-27 is provided in Figure 9. The encoding sequence from nucleotides 37-144 of chKASIII-27 are believed to encode a transit peptide, and the presumed mature protein encoding sequence is from nucleotides 145-1233.

Deduced amino acid sequence of the *C. hookeriana* KAS factor B and KAS factor A cDNA's reveals strong homology to the *Brassica napus* and *Ricinus communis* clones previously reported. The *C. hookeriana* KAS factor B clone is more homologous to the *Ricinus* and *Brassica* KAS factor B clones (94% and 91% respectively) than it is to the *Ricinus* and *Brassica* KAS factor A clones (60% for both). Furthermore, the *C. hookeriana* KAS factor A clone is more homologous to the *Ricinus* and *Brassica* KAS factor A clones (85% and 82% respectively) than it is the *Ricinus* and *Brassica* KAS factor B clone (60% for both). The *C. hookeriana* KAS factor B cDNAs designated as chKAS B-2 and chKAS B-31-7 are 96% identical within the mature portion of the polypeptide. Similarly, the deduced amino acid sequence of the mature protein regions of the *C. hookeriana* KAS factor A clones chKAS A-2-7 and chKAS A-1-6 are 96% identical. The *C. pullcherrima* KAS clones also demonstrate homology to the *R. communis* and *Brassica napus* KAS clones. The mature protein portion of all of the KAS factor A family members in the different *Cuphea* species are 95% identical. Similarly the

mature protein portion of the KAS factor B genes in *Cuphea* are also 95-97% identical with each other. However there is only approximately 60% sequence identity between KAS factor B and KAS factor A clones either within the same or different species of *Cuphea*.

Example 2 Levels and Patterns of Expression

To examine tissue specificity of KAS expression in *Cuphea hookeriana*, Northern blot analysis was conducted using total RNA isolated from seed, root, leaf and flower tissue. Two separate but identical blots were hybridized with either chKAS B-31-7 or chKAS A-2-7 coding region probes. The data from this RNA blot analysis indicate that KAS B is expressed at a similar level in all tissues examined, whereas KAS A expression is detected only in the seed. These results also demonstrate a different level of expression for each of the synthases. KAS A is an abundant message, whereas KAS B is expressed at low levels. Furthermore, even under highly stringent hybridization conditions (65_C, 0.1 X SSC, 0.5% SDS), the KAS A probe hybridizes equally well with two seed transcripts of 2.3 and 1.9 kb. The larger hybridizing band is likely the transcript of the KAS A-2-7 gene since the size of its cDNA is 2046bp, and the number of clones obtained from cDNA screening corresponds well with the apparent mobility of the mRNA and its abundance on the blot.

Example 3 Expression of Plant KAS Genes in E.coli

DNA fragments encoding the mature polypeptide of the *Cuphea hookeriana* KAS A cDNAs and the *Cuphea pullcherrima* KAS B cDNAs were obtained by PCR and cloned into a QIAexpress expression vector (Qiagene). Experimental conditions for maximum level of expression were determined for all of these clones and the parameters for highest level of soluble fraction were identified. Cells are grown in ECLB media containing 1M sorbitol and 2.5 mM betaine overnight and subcultured as a 1:4 dilution in the same medium. Cells are then grown for 2 hours (to approximately .6-.8 O.D.) and induced with 0.4 mM IPTG and allowed to grow for 5 more hours.

Enzyme activity of the affinity purified recombinant enzymes obtained from over-expression of the chKAS A-2-7 and cpuKAS B/8-7A clones was measured using a wide range of acyl-ACP substrates (6:0- to 16:1-ACP). The activity profile for cpuKAS B/8-7A is provided in Fig.10. The data demonstrate that the enzyme is active with all acyl-ACP substrates examined, although activity on 6:0 to 14:0-ACP substrates is substantially greater than the activity on 16:0 and 16:1 substrates.

The activity profile of the *C. hookeriana* KAS A clones chKAS A-2-7 and chKAS A-1-6 is provided in Figure 11. The *C. hookeriana* KAS A clones are most active with C:6, and have the least activity with C:16:0 substrates. However, the activity of this clone on even the preferred C6:0 substrate

is 50 fold lower than the activity of the *C. pullcherrima* KAS B clones.

A fragment containing the mature protein encoding portion of a *R. communis* KAS factor A clone was also cloned into a QIAexpress expression vector, expressed in *E. coli* and the enzyme affinity purified as described above. The activity profile for castor KAS A is provided in Figure 12. Highest activity is observed with C14:0 substrates, although some activity is also seen with C6:0 and C16:1. In comparison, the activity profile obtained from purified *R. communis* KAS factor B also using the QIAexpress expression system is provided in Figure 13. The KAS B clone demonstrates substantially higher levels of activity (10 fold and higher) than the *R. communis* KAS A clone. The preference of the KAS factor B for 6:0- to 14:0-ACP substrates is consistent with the previous observations that this protein provides KAS I activity.

Example 4 KAS and TE Expression in Transgenic Seed

Both the CpFatB1 (*C. hookeriana* thioesterase cDNA; Dehesh *et al.* (1996) *Plant Physiol. 110*:203-210) and the chKAS A-2-7 were PCR amplified, sequenced, and cloned into a napin expression cassette. The napin/cp FatB1 and the napin/KAS A-2-7 fusions were ligated separately into the binary vector pCGN1558 (McBride and Summerfelt (*Pl.Mol.Biol.* (1990) *14*:269-276) and transformed into *A. tumefaciens*, EHA101. The resulting CpFatB1 binary construct is pCGN5400 and the chKAS A-2-7 construct is pCGN5401. *Agrobacterium* mediated transformation of a *Brassica napus* canola variety

was carried out as described by Radke *et al.* (*Theor. Appl. Genet.* (1988) 75:685-694; *Plant Cell Reports* (1992) *11*:499-505). Several transgenic events were produced for each of the pCGN5400 and pCGN5401 constructs.

A double gene construct containing a napin/cpFatB1 expression construct in combination with a napin/chKAS A-2-7 expression construct was also assembled, ligated into a binary vector and used for co-cultivation of a canola *Brassica* variety. The binary construct containing the chFatB1 and chKAS A-2-7 expression constructs is pCGN5413.

Fatty acid analysis of 26 transgenic lines containing chKAS A-2-7 (5401 lines) showed no significant changes in the oil content or profile as compared to similar analyses of wild type canola seeds of the transformed variety.

Fatty acid analysis of 36 transgenic lines containing cpFatB1 (5400 lines) showed increased levels of C:8 and C:10 in transgenic seeds. The highest level of C:8 observed in a pool seed sample was 4.2 mol%. The C:10 levels were between 30 and 35% of the C:8 content. Fatty acid analysis of 25 transgenic lines containing the TE/KAS A tandem (5413 lines) demonstrated an overall increase in both C:8 and C:10 levels relative to those observed with TE containing lines (5400) alone. In lines containing the cpFatB1 construct alone, the average level of C:8 average were 1.5 mol%, whereas the C:8 average levels in TE/KAS A tandem containing lines was 2.37 mol%. The ratio of C:8 to C:10 remained constant in both populations. The number of transgenic events relative to the C:8 content are presented in Figure 14. These data show that the transgenic events with tandem TE/KAS A construct

yield more lines with higher levels of C:8 than those events with single TE construct. For example, several lines containing nearly 7 mole% C8 were obtained with the TE/KAS A pCGN5413 construct, whereas the highest C8 containing line from the pCGN5400 TE alone transformation contained 4.2 mole% C8.

Half seed analysis of the T3 generation of transgenic canola plants expressing a ChFatB2 (*C. hookeriana* thioesterase; Dehesh *et al.* (1996) *The Plant Journal 9*:167-172) indicate that these plant can accumulate up to 22 weight% (33 mol%) of 8:0 and 10:0 fatty acids (4804-22-357). Segregation analysis shows that these transformants contain two loci and that they are now homozygous. Selected plants grown from these half seeds were transferred into the greenhouse and later crossed with T1 transformants that had been transformed with either *Cuphea hookeriana* KAS A (5401) alone or KAS A/CpFatB1 double constructs (5413).

Fatty acid analysis of several events resulting from the crosses between transgenic lines containing ChFatB2 (4804-22-357) and chKAS A-2-7 (5401-9), reveal an increase in the ratio of C:10/C:8 levels (Figure 15). This C:10/C:8 ratio in nearly all of the transgenic events containing ChFatB2 TE alone fluctuates between 3 and 6, whereas in the F1 generation of transgenic containing both the TE and the KAS A-2-7, the ratio can be as high as 22. This increase in C:10 levels is accompanied by an increase in the total C:8 and C:10 content (Figure 16). The sum of the C:8 and C:10 fatty acids in the heterozygous F1 lines is as high as those in the homozygous parent line (4804-22-357), whereas the

heterozygous lines usually contain substantially less C:8 and C:10 than the homozygous lines.

Similar results were observed in F1 generation seeds resulting from crosses performed between 4804-22-357 (ChFatB2) and the 5413-17 event (CpFatB1 and chKAS A-2-7 tandem). Levels of C:8 and C:10 in the 5413-17 line were 6.3 and 2.8 mol% respectively. Data presented in Figure 17 show that there is shift towards C:10 fatty acids as was observed with the 4804-22-357 (ChFatB2) x 5401-9 (chKAS A-2-7) crosses. Furthermore, Figure 18 indicates the presence of two separate populations of heterozygotes. Those containing approximately 9-11 weight percent C:10 + C:8 are believed to represent offspring containing a single copy of the ChFatB1 TE gene and no copies of the CpFatB1 and chKAS A genes from 5413. Those plants containing approximately 15-20 weight percent C:10 + C:8 are believed to represent the heterozygotes containing a single ChFatB1 TE gene as well as the CpFatB1 and chKAS A genes from 5413. Thus, the level of the C:10 + C:8 fatty acids does not decrease to 50% of that detected in parent lines when a copy of the ChKAS A gene is present.

To further characterize the chain length specificity of the *Cuphea hookeriana* KAS A enzyme, crosses between transgenic *Brassica napus* lines containing a California Bay (*Umbellularia californica*) 12:0 specific thioesterase, Uc FatB1 (USPN 5,344,771) and chKAS A-2-7 (5401-9) were made. Half seed analysis of transgenic plants containing Uc fatB1 have previuosly indicated that these plants can accumulate up to 52 mol% C12:0 in the seed oil of homozygous dihaploid

lines (LA86DH186). Crosses between the line LA86DH186 and untransformed control *Brassica* demonstrated a decrease in the C12:0 levels.

However, crosses between LA86DH186 and the 5401-9 hemizygous line led to an accumulation of up to 57 mol% C12:0 in the seed oil of F1 progeny (Figure 19). Interestingly, in crosses with LA86DH186 x untransformed control line and LA86DH186 x 5401-9, levels of C14:0 in the seeds of the F1 progeny decreased to 50% of the levels obtained in homozygous LA86DH186 lines (Figure 20). Furthermore, increases in the proportion of C12:0 fatty acid resulted in a substantial decline in the proportions of all the long-chain fatty acyl groups (C16:0, C18:0, C18:2, and C18:3). These results indicate that the ChKAS A-2-7 is an enzyme with substrate specificity ranging from C6:0 to C10:0-ACP, and that its over-expression ultimately reduces the longer chain acyl-ACP pools.

Further evidence is obtained in support of the chain length specificity of the ChKAS A-2-7 in crosses of the 5401-9 line with a transgenic line (5266) expressing an 18:1/18:0 TE from *Garcinia mangostana* (GarmFatA1, US patent application No. 08/440,845). Transgenic *Brassica* line 5266 has been shown to accumulate up to 24 mol% C18:0 in the seed oil of homozygous lines (Figure 21). However, in the seed oil of F1 progeny of crosses between 5266 and 5401-9 levels of C18:0 were reduced to approximately 12 mol%. Furthermore, levels of C16:0 generated from these crosses was similar to the levels obtained from the seed oil of nontransgenic control plants.

Example 5 ***In vitro*** **Analysis of Plant KAS Enzymes**

Seed extracts were prepared from developing seeds of nontransgenic controls or transgenic *Brassica* expressing chKAS A-2-7 as described in Slabaugh *et al.*(*Plant Journal*, 1998 in press) and Leonard *et al.* (Plant Journal, 1998, in press). *In vitro* fatty acid synthesis assays were performed as described by Post-Beittenmiller (*J. Biol. Chem.* (1991), 266:1858-1865). Extracts were concentrated by ammonium sulfate precipitation and desalting using P-6 columns (Bio-Rad, Hercules, CA). Reactions (65µl) contained 0.1M Tris/HCl (pH 8.0), 1 mM dithiothreitol, 25 mM recombinant spinach ACP1, 1 mM NADH, 2 mM NADPH, 50 µM malonyl-CoA, 10 µM [1-^{14}C]acetyl-CoA (50 mCi/mmol), 1mg/ml BSA, and 0.25 mg/ml seed protein. Selected seed extracts were preincubated with cerulenin at 23°C for 10 min. Reaction products were separated on an 18% acrlamide gel containing 2.25M urea, electroblotted onto to nitrocellulose and quntitated by phosporimaging using Image QuaNT software (Molecular Dynamics, Sunnyvale, CA). Authentic acyl-ACPs were run in parallel, immunoblotted and finally detected by anti-ACP serum to confirm fatty acid chain lengths.

The results (Figure 22) indicate that the fatty acid synthesis capabilities of transgenic *Brasica* (5401-9) seed extracts was greater than that obtained from in the nontransgenic controls as measured by the relative abundance of C8:0- and C10:0-ACP at all time points tested. In addition, pretreatment of the extracts with cerulenin, markedly reduced the synthesis of longer chain fatty acids in both the transgenic and nontransgenic control seed

extracts. However, the extension of the spinach-ACP was much less inhibited in the seed extracts from the transgenic lines than in the seed extracts of nontransgenic control *Brassica*.

These data further support that Ch KAS A-2-7 is a condensing enzyme active on medium chain acyl-ACPs, and that expression of this enzyme in plants results in enlarged substrate pools to be hydrolyzed by medium-chain specific thioesterases. Furthermore, these data suggest that chKAS A-2-7 also is a cerulenin-resistant condensing enzyme.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claim.

MISSING UPON TIME OF PUBLICATION

13. The construct of Claim 5 wherein said encoding sequence is cpuKAS A/p8-9A.

14. The construct of Claim 5 wherein said encoding sequence is chKASIII-27.

15. An improved method for producing medium-chain fatty acids in transgenic plant seeds by expression of a plant medium-chain thioesterase protein heterologous to said transgenic plant,

the improvement comprising expression of a plant synthase factor protein heterologous to said transgenic plant in conjunction with expression of said plant medium-chain thioesterase, whereby the percentage of medium-chain fatty acids produced in seeds expressing both a plant synthase factor protein and a plant medium-chain thioesterase protein is increased as compared to the percentage of medium-chain fatty acids produced in seeds expressing only said plant medium-chain thioesterase protein.

16. The method of Claim 15 wherein said medium-chain thioesterase protein is a ChFatB2 protein.

17. The method of Claim 15 wherein said medium-chain thioesterase protein is a CpFatB1 protein.

18. The method of Claim 15 wherein said medium-chain thioesterase protein is a C12 preferring thioesterase from California bay.

19. The method of Claim 15 wherein said plant synthase factor protein is expressed from a construct according to Claim 1.

20. The method of Claim 19 wherein said synthase factor A protein is from a *Cuphea* species.

21. The method of Claim 20 wherein said *Cuphea* species is *C. hookeriana* or *C. pullcherrima*.

22. A method of altering the medium-chain fatty acid composition in plant seeds expressing a heterologous plant medium-chain preferring thioesterase, wherein said method comprises

providing for expression of a plant synthase factor protein heterologous to said transgenic plant in conjunction with expression of a plant medium-chain thioesterase protein heterologous to said transgenic plant, whereby the composition of medium-chain fatty acids produced in said seeds is modified as compared to the composition of medium-chain fatty acids produced in seeds expressing said plant medium-chain thioesterase protein in the absence of expression of said plant synthase factor protein.

23. The method of Claim 22 wherein said medium-chain thioesterase protein is a ChFatB2 protein.

24. The method of Claim 22 wherein said medium-chain thioesterase protein is a CpFatB1 protein.

25. The method of Claim 22 wherein said medium-chain thioesterase protein is a C12 preferring thioesterase from California bay.

26. The method of Claim 22 wherein said plant synthase factor protein is expressed from a construct according to Claim 1.

27. The method of Claim 26 wherein said synthase factor A protein is from a *Cuphea* species.

28. The method of Claim 27 wherein said *Cuphea* species is *C. hookeriana* or *C. pullcherrima*.

29. The method of Claim 22 wherein said fatty acid composition is enriched for C10 fatty acids.

30. The method of Claim 22 wherein said fatty acid composition is enriched for C12 fatty acids.

31. The method of Claim 22 wherein said fatty acid composition is enriched for at least one medium chain fatty acid and at least one other medium chain fatty acid is decreased.

32. The method of Claim 31 wherein said enriched fatty acid is C12 and said decreased fatty acid is C14.

y66

```
AGC TCC ACC GCG GTG GCG GCC GCT CTA GAA CTA GTG GAT CCC CCG GGC     48
Ser Ser Thr Ala Val Ala Ala Ala Leu Glu Leu Val Asp Pro Pro Gly

TGC AGG AAT TCG GCA CGA GCC GAT CTC GGT GCC GAC CGC CTC TCC AAG     96
Cys Arg Asn Ser Ala Arg Ala Asp Leu Gly Ala Asp Arg Leu Ser Lys

ATC GAC AAG GAG AGA GCC GGA GTG CTG GTC GGA ACA GGA ATG GGT GGT    144
Ile Asp Lys Glu Arg Ala Gly Val Leu Val Gly Thr Gly Met Gly Gly

CTG ACT GTC TTC TCT GAC GGG GTT CAG TCT CTT ATC GAG AAG GGT CAC    192
Leu Thr Val Phe Ser Asp Gly Val Gln Ser Leu Ile Glu Lys Gly His

CGG AAA ATC ACC CCT TTC TTC ATC CCC TAT GCC ATT ACA AAC ATG GGG    240
Arg Lys Ile Thr Pro Phe Phe Ile Pro Tyr Ala Ile Thr Asn Met Gly

TCT GCC CTG CTC GCT ATC GAA TTT GGT CTC ATG GGC CCA AAC TAT TCA    288
Ser Ala Leu Leu Ala Ile Glu Phe Gly Leu Met Gly Pro Asn Tyr Ser

ATT TCC ACT GCA TGT GCC ACT TCC AAC TAC TGC TTC CAT GCT GCC GCT    336
Ile Ser Thr Ala Cys Ala Thr Ser Asn Tyr Cys Phe His Ala Ala Ala

AAT CAT ATC CGC CGT GGT GAG GCT GAT CTT ATG ATT GCT GGA GGC ACT    384
Asn His Ile Arg Arg Gly Glu Ala Asp Leu Met Ile Ala Gly Gly Thr
```

FIGURE 1
1 OF 4

```
GAG GCC GCA ATC ATT CCA ATT GGG TTG GGA GGC TTT GTG GCT TGC AGG    432
Glu Ala Ala Ile Ile Pro Ile Gly Leu Gly Gly Phe Val Ala Cys Arg

GCT TTG TCT CAA AGG AAC GAT GAC CCG CAG ACT GCC TCT AGG CCC TGG    480
Ala Leu Ser Gln Arg Asn Asp Asp Pro Gln Thr Ala Ser Arg Pro Trp

GAT AAA GAC CGT GAT GGT TTT GTG ATG GGT GAA GGT GCT GGA GTG TTG    528
Asp Lys Asp Arg Asp Gly Phe Val Met Gly Glu Gly Ala Gly Val Leu

GTG ATG GAG AGC TTG GAA CAT GCA ATG AGA CGA GGA GCA CCG ATT ATT    576
Val Met Glu Ser Leu Glu His Ala Met Arg Arg Gly Ala Pro Ile Ile

GCA GAG TAT TTG GGA GGT GCA ATC AAC TGT GAT GCT TAT CAC ATG ACT    624
Ala Glu Tyr Leu Gly Gly Ala Ile Asn Cys Asp Ala Tyr His Met Thr

GAT CCA AGG GCT GAT GGT CTT GGT GTC TCT TCT TGC ATT GAG AGT AGC    672
Asp Pro Arg Ala Asp Gly Leu Gly Val Ser Ser Cys Ile Glu Ser Ser

CTT GAA GAT GCT GGC GTC TCA CCT GAA GAG GTC AAT TAC ATA AAT GCT    720
Leu Glu Asp Ala Gly Val Ser Pro Glu Glu Val Asn Tyr Ile Asn Ala

CAT GCG ACT TCT ACT CTA GCT GGG GAT CTC GCC GAG ATA AAT GCC ATC    768
His Ala Thr Ser Thr Leu Ala Gly Asp Leu Ala Glu Ile Asn Ala Ile
```

FIGURE 1
2 OF 4

```
AAG AAG GTT TTC AAG AAC ACA AAG GAT ATC AAA ATT AAT GCA ACT AAG      816
Lys Lys Val Phe Lys Asn Thr Lys Asp Ile Lys Ile Asn Ala Thr Lys

TCA ATG ATC GGA CAC TGT CTT GGA GCA TCT GGA GGT CTT GAA GCT ATA      864
Ser Met Ile Gly His Cys Leu Gly Ala Ser Gly Gly Leu Glu Ala Ile

GCG ACT ATT AAG GGA ATA AAC ACC GGC TGG CTT CAT CCC AGC ATT AAT      912
Ala Thr Ile Lys Gly Ile Asn Thr Gly Trp Leu His Pro Ser Ile Asn

CAA TTC AAT CCT GAG CCA TCG GTG GAG TTC GAC ACT GTT GCC AAC AAG      960
Gln Phe Asn Pro Glu Pro Ser Val Glu Phe Asp Thr Val Ala Asn Lys

AAG CAG CAA CAC GAA GTT AAC GTT GCG ATC TCG AAT TCA TTC GGA TTT     1008
Lys Gln Gln His Glu Val Asn Val Ala Ile Ser Asn Ser Phe Gly Phe

GGA GGC CAC AAC TCA GTC GTG GCT TTC TCG GCT TTC AAG CCA TGATTA      1056
Gly Gly His Asn Ser Val Val Ala Phe Ser Ala Phe Lys Pro

CCCATTTCAC AAGGTACTTG TCATTGAGAA TACGGATTAT GGACTTGCAG AGTAATTTCC   1116

CCATGTTTGT CGGAAGAGCA TATTACCACG GTTGTCCGTC AAACCCATTT AGGATACTGT   1176
```

FIGURE 1
3 OF 4

```
TCTATGTAAT AAAACTAAGG ATTATTAATT TCCCTTTTAA TCCTGTCTCC AGTTTGAGCA   1236
TGAAATTATA TTTATTTTAT CTTAGAAAGG TCAAATAAGA TTTTGTTTTA CCTCTGTAAA   1296
ACTTTTGTTT GTATTGGAAA GGAAGTGCCG TCTCAAAAAA AAAAAAAAAA AA            1348
```

FIGURE 1
4 OF 4

```
Sequence Range: 1 to 1704

                        10          20          30          40
AAA TTA ACC CTC ACT AAA GGG AAC AAA AGC TGG AGC TCC ACC GNG GTG
Lys Leu Thr Leu Thr Lys Gly Asn Lys Ser Trp Ser Ser Thr Xxx Val>

50          60          70          80          90
             *
GCG GCC GCT CTA GAA CTA GTG GAT CCC CCG GGC TGC AGG AAT TCG GCA
Ala Ala Ala Leu Glu Leu Val Asp Pro Pro Gly Cys Arg Asn Ser Ala>

  100         110         120         130         140
                            *
CGA GCC GGC ATG GGC CTC GTC TCC GTA TTC GGC TCC GAC GTC GAC TCT
Arg Ala Gly Met Gly Leu Val Ser Val Phe Gly Ser Asp Val Asp Ser>

    150         160         170         180         190
                                         *
TAT TAC GAA AAG CTC CTC TCC GGC GAG AGC GGG ATC AGC TTA ATC GAC
Tyr Tyr Glu Lys Leu Leu Ser Gly Glu Ser Gly Ile Ser Leu Ile Asp>

       200         210         220         230         240
                                                         *
CGC TTC GAC GCT TCC AAG TTC CCC ACC AGG TTC GGC GGC CAG ATC CGG
Arg Phe Asp Ala Ser Lys Phe Pro Thr Arg Phe Gly Gly Gln Ile Arg>

          250         260         270         280
GGA TTC AAC GCG ACG GGA TAC ATC GAC GGG AAG AAC GAC AGG AGG CTC
Gly Phe Asn Ala Thr Gly Tyr Ile Asp Gly Lys Asn Asp Arg Arg Leu>

90          300         310         320         330
             *
GAC GAT TGC CTC CGC TAC TGC ATT GTC GCC GGG AAG AAG GCT CTC GAA
Asp Asp Cys Leu Arg Tyr Cys Ile Val Ala Gly Lys Lys Ala Leu Glu>
```

FIGURE 2
1/5

```
 340         350         360         370         380
                           *
AAT TCC GAT CTC GGC GGT GAA AGC CTC TCC AAG ATT GAT AAG GAG AGA
Asn Ser Asp Leu Gly Gly Glu Ser Leu Ser Lys Ile Asp Lys Glu Arg>

    390         400         410         420         430
                                          *
GCT GGA GTG CTA GTT GGA ACT GGT ATG GGT GGC CTA ACC GTC TTC TCT
Ala Gly Val Leu Val Gly Thr Gly Met Gly Gly Leu Thr Val Phe Ser>

       440         450         460         470         480
                                                         *
GAC GGG GTT CAG AAT CTC ATC GAG AAA GGT CAC CGG AAG ATC TCC CCG
Asp Gly Val Gln Asn Leu Ile Glu Lys Gly His Arg Lys Ile Ser Pro>

          490         500         510         520
TTT TTC ATT CCC TAT GCC ATT ACA AAC ATG GGG TCT GCT CTG CTT GCC
Phe Phe Ile Pro Tyr Ala Ile Thr Asn Met Gly Ser Ala Leu Leu Ala>

30          540         550         560         570
             *
ATC GAT TTG GGT CTG ATG GGC CCA AAC TAT TCG ATT TCA ACT GCA TGT
Ile Asp Leu Gly Leu Met Gly Pro Asn Tyr Ser Ile Ser Thr Ala Cys>

  580         590         600         610         620
                            *
GCT ACT TCC AAC TAC TGC TTT TAT GCC GCT GCC AAT CAT ATC CGC CGA
Ala Thr Ser Asn Tyr Cys Phe Tyr Ala Ala Ala Asn His Ile Arg Arg>

    630         640         650         660         670
                                          *
GGC GAG GCT GAC CTC ATG ATT GCT GGA GGA ACT GAG GCT GCA ATC ATT
Gly Glu Ala Asp Leu Met Ile Ala Gly Gly Thr Glu Ala Ala Ile Ile>
```

FIGURE 2
2/5

```
        680         690         700         710         720
                                                           *
CCA ATT GGG TTA GGA GGA TTC GTT GCC TGC AGG GCT TTA TCT CAA AGG
Pro Ile Gly Leu Gly Gly Phe Val Ala Cys Arg Ala Leu Ser Gln Arg>

           730         740         750         760
AAT GAT GAC CCT CAG ACT GCC TCA AGG CCG TGG GAT AAG GAC CGT GAT
Asn Asp Asp Pro Gln Thr Ala Ser Arg Pro Trp Asp Lys Asp Arg Asp>

70          780         790         800         810
              *
GGT TTT GTG ATG GGC GAA GGG GCT GGA GTA TTG GTT ATG GAG AGC TTG
Gly Phe Val Met Gly Glu Gly Ala Gly Val Leu Val Met Glu Ser Leu>

  820         830         840         850         860
                            *
GAA CAT GCA ATG AAA CGA GGA GCG CCG ATT ATT GCA GAA TAT TTG GGA
Glu His Ala Met Lys Arg Gly Ala Pro Ile Ile Ala Glu Tyr Leu Gly>

    870         880         890         900         910
                                          *
GGT GCA GTC AAT TGT GAT GCT TAT CAT ATG ACT GAT CCA AGG GCT GAT
Gly Ala Val Asn Cys Asp Ala Tyr His Met Thr Asp Pro Arg Ala Asp>

       920         930         940         950         960
                                                         *
GGG CTT GGT GTC TCC TCT TGC ATT GAG AGC AGT CTG GAA GAT GCT GGG
Gly Leu Gly Val Ser Ser Cys Ile Glu Ser Ser Leu Glu Asp Ala Gly>

          970         980         990        1000
GTC TCA CCT GAA GAG GTC AAT TAC ATA AAT GCT CAT GCG ACT TCC ACT
Val Ser Pro Glu Glu Val Asn Tyr Ile Asn Ala His Ala Thr Ser Thr>
```

FIGURE 2
3/5

```
10          1020        1030        1040        1050
              *
CTT GCT GGG GAT CTT GCC GAG ATA AAT GCC ATC AAG AAG GTT TTC AAG
Leu Ala Gly Asp Leu Ala Glu Ile Asn Ala Ile Lys Lys Val Phe Lys>

 1060        1070        1080        1090        1100
                           *
AAC ACC AAG GAA ATC ACA ATC AAT GCA ACT AAG TCG ATG ATC GGA CAC
Asn Thr Lys Glu Ile Thr Ile Asn Ala Thr Lys Ser Met Ile Gly His>

   1110        1120        1130        1140        1150
                                         *
TGT CTT GGA GCA TCA GGG GGT CTT GAA GCC ATT GCG ACA ATT AAG GGA
Cys Leu Gly Ala Ser Gly Gly Leu Glu Ala Ile Ala Thr Ile Lys Gly>

       1160        1170        1180        1190        1200
                                                         *
ATA ACC ACC GGC TGG CTT CAT CCC AGC ATA AAC CAA TTC AAT CCC GAG
Ile Thr Thr Gly Trp Leu His Pro Ser Ile Asn Gln Phe Asn Pro Glu>

          1210        1220        1230        1240
CCA TCA GTG GAA TTC GAC ACA GTT GCC AAC AAG AAG CAG CAA CAT GAA
Pro Ser Val Glu Phe Asp Thr Val Ala Asn Lys Lys Gln Gln His Glu>

50          1260        1270        1280        1290
              *
GTG AAT GTT GCT ATC TCA AAT TCA TTC GGA TTC GGA GGC CAC AAC TCA
Val Asn Val Ala Ile Ser Asn Ser Phe Gly Phe Gly Gly His Asn Ser>

 1300        1310        1320        1330        1340
                           *
GTT GTA GCT TTC TCA GCC TTC AAG CCA TGA TTA CTC GGT TCA AAT GCA
Val Val Ala Phe Ser Ala Phe Lys Pro
```

FIGURE 2
4/5

AATTTGTTGC TGAGACAGTG AGCTTCAACT TGCAGAGCAA TTTTTTACAT GCCTTGTCGT

CGGAAGAGCG TAATACCGGG ATAGTTCCTT GATAGTTCAT TTAGGATGTT TTACTGCAAT

AATCGAAGAT TATTTCCATT CTAATCCAGT CTCCGNCGAG TTTGAGAATC TATCTGTTTG

TATTAGAAAG AACGAGGCAA GATTTTGTTT CATGTTTGTG TTTGTATTAC TTTCTTTTTG

CCCTTGTCAA TGGCATTTAA GATAAGCTTA TAAAAAAAAA AAAAAAAAAA AAAACTCGAG

GGGGGGCCCG GTACCCAATT CGCCCTATAG TGAGTCGTAT GACAATTCAC TGTCCGTCGG

FIGURE 2
5/5

```
        10         20         30         40         50         60
                                                                 *
ACTAAAGGGA ACAAAAGCTG GAGCTCCACC GCGGTGGCGG CCGCTCTAGA ACTAGTGGAT

        70         80         90        100        110        120
                                                                 *
CCCCCGGGCT GCAGGAATTC GGCACGAGTT TTCTTACTTG GGTCGGCTCA GCTCAGGTGT

         130          140          150          160
TCCA ATG GCG ACC GCT TCT TGC ATG GTT GCG TCC CCT TTC TGT ACG TGG
     Met Ala Thr Ala Ser Cys Met Val Ala Ser Pro Phe Cys Thr Trp

170          180          190          200          210
               *
CTC GTA GCT GCA TGC ATG CCC ACT TCA TCC GAC AAC GAC CCA CGT TCC
Leu Val Ala Ala Cys Met Pro Thr Ser Ser Asp Asn Asp Pro Arg Ser

220          230          240          250          260
                            *
CTT TCC CAC AAG CGG CTC CGC CTC TCC CGT CGC CGG AGG ACT CTC TCC
Leu Ser His Lys Arg Leu Arg Leu Ser Arg Arg Arg Arg Thr Leu Ser

   270          280          290          300          310
                                            *
TCC CAT TGC TCC CTC CGC GGA TCC ACC TTC CAA TGC CTC GAT CCT TGC
Ser His Cys Ser Leu Arg Gly Ser Thr Phe Gln Cys Leu Asp Pro Cys

      320          330          340          350          360
                                                            *
AAC CAG CAA CGC TTC CTC GGG GAT AAC GGA TTC GCT TCC CTC TTC GGA
Asn Gln Gln Arg Phe Leu Gly Asp Asn Gly Phe Ala Ser Leu Phe Gly
```

FIGURE 3 1/6

```
          370            380            390            400
TCC AAG CCT CTT CGT TCA AAT CGC GGC CAC CTG AGG CTC GGC CGC ACT
Ser Lys Pro Leu Arg Ser Asn Arg Gly His Leu Arg Leu Gly Arg Thr

410           420           430            440           450
               *
TCC CAT TCC GGG GAG GTC ATG GCT GTG GCT ATG CAA CCT GCA CAG GAA
Ser His Ser Gly Glu Val Met Ala Val Ala Met Gln Pro Ala Gln Glu

460            470           480           490            500
                               *
GTC TCC ACA AAT AAG AAA CCT GCT ACC AAG CAA AGG CGA GTA GTT GTG
Val Ser Thr Asn Lys Lys Pro Ala Thr Lys Gln Arg Arg Val Val Val

   510           520            530           540           550
                                               *
ACA GGT ATG GGC GTG GTG ACT CCT CTA GGC CAT GAC CCC GAT GTT TAC
Thr Gly Met Gly Val Val Thr Pro Leu Gly His Asp Pro Asp Val Tyr

      560           570           580            590           600
                                                                 *
TAC AAC AAT CTC CTA GAC GGA ATA AGT GGC ATA AGT GAG ATA GAG AAC
Tyr Asn Asn Leu Leu Asp Gly Ile Ser Gly Ile Ser Glu Ile Glu Asn

          610            620           630           640
TTC GAC TGC TCT CAG TTT CCC ACG AGA ATT GCC GGA GAG ATC AAG TCT
Phe Asp Cys Ser Gln Phe Pro Thr Arg Ile Ala Gly Glu Ile Lys Ser

650           660           670            680           690
               *
TTT TCC ACA GAT GGC TGG GTG GCC CCA AAG TTC TCC GAG AGG ATG GAC
Phe Ser Thr Asp Gly Trp Val Ala Pro Lys Phe Ser Glu Arg Met Asp
```

FIGURE 3 2 OF 6

12/66

```
  700             710         720         730          740
                      *
AAG TTC ATG CTT TAC ATG CTG ACT GCA GGC AAG AAA GCA TTA GCA GAT
Lys Phe Met Leu Tyr Met Leu Thr Ala Gly Lys Lys Ala Leu Ala Asp

  750          760         770          780         790
                                         *
GGT GGA ATC ACT GAA GAT GCG ATG AAA GAG CTC AAT AAA AGA AAG TGT
Gly Gly Ile Thr Glu Asp Ala Met Lys Glu Leu Asn Lys Arg Lys Cys

     800          810         820         830         840
                                                        *
GGA GTT CTC ATT GGC TCC GGA TTG GGC GGT ATG AAG GTA TTC AGC GAT
Gly Val Leu Ile Gly Ser Gly Leu Gly Gly Met Lys Val Phe Ser Asp

       850         860         870         880
TCC ATT GAA GCT CTG AGG ACT TCA TAT AAG AAG ATC AGT CCC TTT TGT
Ser Ile Glu Ala Leu Arg Thr Ser Tyr Lys Lys Ile Ser Pro Phe Cys

890        900         910         920          930
            *
GTA CCT TTT TCT ACC ACA AAT ATG GGA TCC GCT ATT CTT GCA ATG GAC
Val Pro Phe Ser Thr Thr Asn Met Gly Ser Ala Ile Leu Ala Met Asp

940         950         960         970         980
                          *
TTG GGA TGG ATG GGC CCT AAC TAT TCG ATA TCA ACT GCC TGT GCA ACA
Leu Gly Trp Met Gly Pro Asn Tyr Ser Ile Ser Thr Ala Cys Ala Thr

  990         1000        1010         1020        1030
                                         *
AGT AAC TTC TGT ATA CTG AAT GCT GCG AAC CAC ATA ATC AAA GGC GAA
Ser Asn Phe Cys Ile Leu Asn Ala Ala Asn His Ile Ile Lys Gly Glu
```

FIGURE 3 3 OF 6

```
       1040            1050            1060             1070            1080
          *
GCA GAC ATG ATG CTT TGT GGT GGC TCG GAT GCG GCC GTT TTA CCT GTT
Ala Asp Met Met Leu Cys Gly Gly Ser Asp Ala Ala Val Leu Pro Val

          1090             1100            1110            1120
GGT TTG GGA GGT TTC GTA GCA TGC CGA GCT TTG TCA CAG AGG AAT AAT
Gly Leu Gly Gly Phe Val Ala Cys Arg Ala Leu Ser Gln Arg Asn Asn

1130           1140            1150             1160            1170
                  *
GAC CCT ACC AAA GCT TCG AGA CCA TGG GAC AGT AAT CGT GAT GGA TTT
Asp Pro Thr Lys Ala Ser Arg Pro Trp Asp Ser Asn Arg Asp Gly Phe

  1180             1190            1200            1210             1220
                                      *
GTG ATG GGA GAA GGA GCT GGA GTT TTA CTT CTT GAG GAG TTA GAG CAT
Val Met Gly Glu Gly Ala Gly Val Leu Leu Leu Glu Glu Leu Glu His

     1230            1240             1250            1260            1270
                                                         *
GCA AAG AAA AGA GGT GCA ACC ATT TAT GCG GAA TTT CTA GGT GGG AGT
Ala Lys Lys Arg Gly Ala Thr Ile Tyr Ala Glu Phe Leu Gly Gly Ser

        1280            1290            1300             1310            1320
                                                                            *
TTC ACT TGC GAC GCC TAC CAC ATG ACC GAG CCT CAC CCT GAA GGA GCT
Phe Thr Cys Asp Ala Tyr His Met Thr Glu Pro His Pro Glu Gly Ala

          1330             1340            1350            1360
GGT GTG ATC CTC TGC ATA GAG AAG GCC TTG GCT CAG TCC GGA GTC TCG
Gly Val Ile Leu Cys Ile Glu Lys Ala Leu Ala Gln Ser Gly Val Ser
```

FIGURE 3 4 OF 6

```
1370        1380        1390        1400        1410
              *
  AGG GAA GAC GTA AAT TAC ATA AAT GCG CAT GCA ACT TCC ACT CCT GCT
  Arg Glu Asp Val Asn Tyr Ile Asn Ala His Ala Thr Ser Thr Pro Ala

 1420        1430        1440        1450        1460
                           *
  GGA GAT ATC AAG GAA TAC CAA GCT CTC GCC CAC TGT TTC GGC CAA AAC
  Gly Asp Ile Lys Glu Tyr Gln Ala Leu Ala His Cys Phe Gly Gln Asn

   1470        1480        1490        1500        1510
                                         *
  AGT GAG CTG AGA GTG AAT TCC ACC AAA TCG ATG ATC GGT CAC CTT CTT
  Ser Glu Leu Arg Val Asn Ser Thr Lys Ser Met Ile Gly His Leu Leu

     1520        1530        1540        1550        1560
                                                       *
  GGA GGA GCT GGT GGC GTA GAA GCA GTT GCA GTA GTT CAG GCA ATA AGG
  Gly Gly Ala Gly Gly Val Glu Ala Val Ala Val Val Gln Ala Ile Arg

       1570        1580        1590        1600
  ACA GGA TGG ATC CAT CCA AAT ATT AAT TTG GAA GAC CCG GAC GAA GGC
  Thr Gly Trp Ile His Pro Asn Ile Asn Leu Glu Asp Pro Asp Glu Gly

1610        1620        1630        1640        1650
              *
  GTG GAT GCA AAA CTG CTC GTC GGC CCT AAG AAG GAG AAA CTG AAG GTC
  Val Asp Ala Lys Leu Leu Val Gly Pro Lys Lys Glu Lys Leu Lys Val

 1660        1670        1680        1690        1700
                           *
  AAG GTC GGT TTG TCC AAT TCA TTT GGG TTC GGC GGC CAT AAC TCA TCC
  Lys Val Gly Leu Ser Asn Ser Phe Gly Phe Gly Gly His Asn Ser Ser
```

FIGURE 3 5 OF 6

```
 1710        1720         1730      1740      1750      1760
                                       *
ATA CTA TTT GCC CCC TGC AAC TAG A AAAGAGTCTG TGGAAGCCGA GAGTCTTTGA
Ile Leu Phe Ala Pro Cys Asn ***

      1770       1780       1790       1800       1810       1820
                                          *
GAACTCATGC ACGTTAGTAG CTTCTTATGC CTCTGAAACC GAGATAGACC GGCTACTCGA

      1830       1840       1850       1860       1870       1880
                                          *
GGGGATGCCA AAGATACTCC TTGCCGGTAT TGGTGTTAAG AGATCACTGC TTGTCCCTTT

      1890       1900       1910       1920       1930       1940
                                          *
TATTTTCTTC TTCTTTTGAG AGCTTTAACC GAGGTAGTCG TATTTTCGAG CTTTTCGAAT

      1950       1960       1970       1980       1990       2000
                                          *
ACATGTTCGT TATCGGATCA ATGTGTTTCT TCTAAGATCA TTTGTAATGC ATATTTTGAA

      2010       2020       2030       2040
                                          *
AAACCACATC TCAGTATGCA AAATAAAAAA AAAAAAAAAA AAAAAA
```

FIGURE 3 6 OF 6

Sequence Range: 1 to 1921

```
        10         20         30         40         50         60
                                                                 *
CGGCACGAGG TCACCTCTTA CCTCGCCTGC TTCGAGCCCT GCCATGACTA CTACACCTCC

        70         80         90        100        110        120
                                                                 *
GCATCCTTGT TCGGATCCAG GCCCATCCGC ACCACCCGCA GGCACCGGAG GCTCAATCGA

       130        140        150        160        170        180
                                                                 *
GCTTCCCCTT CCGGGGAGGC AATGGCTGTG GCTCTGCAAC CTGCACAGGA AGTTACCACA

         190             200             210             220
AAG AAG AAG CCA AGT ATC AAA CAG CGG CGA GTA GTT GTG ACT GGA ATG
Lys Lys Lys Pro Ser Ile Lys Gln Arg Arg Val Val Val Thr Gly Met>

230             240             250             260             270
                  *
GGT GTG GTG ACT CCT CTA GGC CAT GAC CCT GAT GTT TTC TAC AAT AAT
Gly Val Val Thr Pro Leu Gly His Asp Pro Asp Val Phe Tyr Asn Asn>

    280             290             300             310             320
                                      *
CTG CTT GAT GGA ACG AGT GGC ATA AGT GAG ATA GAG ACC TTT GAT TGT
Leu Leu Asp Gly Thr Ser Gly Ile Ser Glu Ile Glu Thr Phe Asp Cys>

      330             340             350             360             370
                                                        *
GCT CAA TTT CCT ACG AGA ATT GCT GGA GAG ATC AAG TCT TTC TCC ACA
Ala Gln Phe Pro Thr Arg Ile Ala Gly Glu Ile Lys Ser Phe Ser Thr>
```

FIGURE 4
1/6

```
        380         390          400         410         420
                                                           *
GAT GGT TGG GTG GCC CCG AAG CTC TCC AAG AGG ATG GAC AAG TTC ATG
Asp Gly Trp Val Ala Pro Lys Leu Ser Lys Arg Met Asp Lys Phe Met>

           430         440         450          460
CTT TAC ATG CTG ACT GCC GGC AAG AAA GCA TTA ACA AAT GGT GGA ATC
Leu Tyr Met Leu Thr Ala Gly Lys Lys Ala Leu Thr Asn Gly Gly Ile>

470         480          490         500         510
              *
ACC GAA GAT GTG ATG AAA GAG CTA GAT AAA AGA AAA TGC GGA GTT CTC
Thr Glu Asp Val Met Lys Glu Leu Asp Lys Arg Lys Cys Gly Val Leu>

   520         530         540          550         560
                             *
ATT GGC TCA GCA ATG GGT GGA ATG AAG GTA TTC AAT GAT GCC ATT GAA
Ile Gly Ser Ala Met Gly Gly Met Lys Val Phe Asn Asp Ala Ile Glu>

     570          580         590         600          610
                                            *
GCC CTA AGG ATT TCA TAT AAG AAG ATG AAT CCC TTT TGT GTA CCT TTC
Ala Leu Arg Ile Ser Tyr Lys Lys Met Asn Pro Phe Cys Val Pro Phe>

        620         630          640         650         660
                                                           *
GCT ACC ACA AAT ATG GGA TCA GCT ATG CTT GCA ATG GAC TTG GGA TGG
Ala Thr Thr Asn Met Gly Ser Ala Met Leu Ala Met Asp Leu Gly Trp>

           670         680         690          700
ATG GGC CCC AAC TAC TCG ATA TCT ACT GCT TGT GCA ACG AGT AAC TTT
Met Gly Pro Asn Tyr Ser Ile Ser Thr Ala Cys Ala Thr Ser Asn Phe>
```

FIGURE 4

2/6

```
710         720        730        740        750
             *
TGT ATC CTG AAT GCT GCG AAC CAC ATA ATC AGA GGC GAA GCA GAT GTG
Cys Ile Leu Asn Ala Ala Asn His Ile Ile Arg Gly Glu Ala Asp Val>

  760        770        780        790        800
                         *
ATG CTT TGC GGG GGC TCA GAT GCG GTA ATC ATA CCT ATT GGT ATG GGA
Met Leu Cys Gly Gly Ser Asp Ala Val Ile Ile Pro Ile Gly Met Gly>

    810        820        830        840        850
                                      *
GGT TTT GTT GCA TGC CGA GCT TTG TCA CAG AGA AAT GCC GAC CCT ACT
Gly Phe Val Ala Cys Arg Ala Leu Ser Gln Arg Asn Ala Asp Pro Thr>

       860        870        880        890        900
                                                    *
AAA GCT TCA AGA CCA TGG GAC AGT AAT CGT GAT GGA TTT GTT ATG GGG
Lys Ala Ser Arg Pro Trp Asp Ser Asn Arg Asp Gly Phe Val Met Gly>

         910        920        930        940
GAA GGA GCT GGA GTG CTA CTA CTA GAG GAG TTA GAG CAT GCA AAG AAA
Glu Gly Ala Gly Val Leu Leu Leu Glu Glu Leu Glu His Ala Lys Lys>

950         960        970        980        990
             *
AGA GGT GCG ACT ATT TAC GCA GAA TTT CTA GGT GGA AGT TTC ACT TGC
Arg Gly Ala Thr Ile Tyr Ala Glu Phe Leu Gly Gly Ser Phe Thr Cys>

 1000       1010       1020       1030       1040
                         *
GAT GCC TAC CAC ATG ACC GAG CCT CAC CCT GAT GGA GCT GGA GTG ATT
Asp Ala Tyr His Met Thr Glu Pro His Pro Asp Gly Ala Gly Val Ile>
```

FIGURE 4
3/6

```
   1050          1060          1070          1080          1090
                                               *
CTC TGC ATA GAG AAG GCT TTG GCT CAG TCA GGA GTC TCT AGG GAA GAC
Leu Cys Ile Glu Lys Ala Leu Ala Gln Ser Gly Val Ser Arg Glu Asp>

      1100          1110          1120          1130          1140
                                                                 *
GTA AAT TAC ATA AAT GCA CAT GCC ACA TCC ACT CCA GCT GGA GAT ATC
Val Asn Tyr Ile Asn Ala His Ala Thr Ser Thr Pro Ala Gly Asp Ile>

          1150          1160          1170          1180
AAA GAG TAC CAA GCT CTT ATC CAC TGT TTC GGC CAA AAC AAC GAG TTA
Lys Glu Tyr Gln Ala Leu Ile His Cys Phe Gly Gln Asn Asn Glu Leu>

1190          1200          1210          1220          1230
                *
AAA GTG AAT TCT ACC AAA TCA ATG ATT GGT CAC CTT CTC GGA GCA GCC
Lys Val Asn Ser Thr Lys Ser Met Ile Gly His Leu Leu Gly Ala Ala>

  1240          1250          1260          1270          1280
                                *
GGT GGT GTG GAA GCA GTT TCA GTA GTT CAG GCA ATA AGG ACT GGG TGG
Gly Gly Val Glu Ala Val Ser Val Val Gln Ala Ile Arg Thr Gly Trp>

   1290          1300          1310          1320          1330
                                               *
ATC CAT CCG AAT ATT AAT TTG GAA AAC CCA GAT GAA GGC GTG GAT ACC
Ile His Pro Asn Ile Asn Leu Glu Asn Pro Asp Glu Gly Val Asp Thr>

      1340          1350          1360          1370          1380
                                                                 *
AAA TTG CTC GTG GGC CCT AAG AAG GAG AGA CTG AAC ATT AAG GTC GGT
Lys Leu Leu Val Gly Pro Lys Lys Glu Arg Leu Asn Ile Lys Val Gly>
```

FIGURE 4
4/6

```
       1390        1400        1410        1420
TTG TCT AAT TCA TTC GGG TTT GGT GGG CAC AAC TCG TCC ATA CTC TTC
Leu Ser Asn Ser Phe Gly Phe Gly Gly His Asn Ser Ser Ile Leu Phe>

1430        1440       1450       1460       1470       1480
               *
GCC CCT TAC AAC TAG GGCGTTT CATGTGTGGA ATTCTACTCA ATCTATCAAA
Ala Pro Tyr Asn ***>

      1490       1500       1510       1520       1530       1540
                    *
GCTGAAGTTT TGAGGACTCC AGCATGTTGG TAGCTCCTTA CGTCTCTAGA CATGCCCATG

      1550       1560       1570       1580       1590       1600
                    *
AGTTTTGTGT CGGGAGCTGT AGTCGGAACC ATGACGGATT GAGTACTCAT GGCGACACAG

      1610       1620       1630       1640       1650       1660
                    *
GATATACTCC TTGCTAGAAT TGTTAGAGCA CTATTCATTA TCCCATTTTT TTTCTGAAAT

      1670       1680       1690       1700       1710       1720
                    *
CTCCCTCCTT ACGGTAGTTG TACTTTCGAG CGTTTCATCG AGTCAGTGAA GAAGAGAACA

      1730       1740       1750       1760       1770       1780
                    *
AAGCTAACTC GGGCACGTAG TAACCATTTG CCCTTTGTTT TGCTCTCTAT TTTATCGCCG

      1790       1800       1810       1820       1830       1840
                    *
TTTTGTGGGT TAAAATTTGT AAAACTAGAC GACTGGTTTG TTTTCTCTTG ATCATTGGAG
```

FIGURE 4
5/6

```
      1850       1860       1870       1880       1890       1900
                    *
ATGTATGGCC ATATTTGCCT TTCATTGATG ATAAAAAAAA AAAAAAAAAA AAAAAAAAAA

      1910       1920
                    *
AAAAAAAAAA AAAAAAAAAA A
```

FIGURE 4
6/6

```
CTGGTACGCC TGCAGGTACC GGTCCGGAAT TCCCGGGTCG ACCCACGCGT CCGTCTTCCC    60

ACTCCGATCG TTCTTCTTCC ACCGCATCTC TTCTCTTCTC TTGGCTTCTC CGCCATCCTC   120

CGCCGCC ATG CAT TCC CTC CAG TCA CCC TCC CTT CGG GCC TCC CCG CTC     169
        Met His Ser Leu Gln Ser Pro Ser Leu Arg Ala Ser Pro Leu
        1               5                   10

GAC CCC TTC CGC CCC AAA TCA TCC ACC GTC CGC CCC CTC CAC CGA GCA     217
Asp Pro Phe Arg Pro Lys Ser Ser Thr Val Arg Pro Leu His Arg Ala
15                  20                  25                  30

TCA ATT CCC AAC GTC CGG GCC GCT TCC CCC ACC GTC TCC GCT CCC AAG     265
Ser Ile Pro Asn Val Arg Ala Ala Ser Pro Thr Val Ser Ala Pro Lys
                35                  40                  45

CGC GAG ACC GAC CCC AAG AAG CGC GTC GTG ATC ACC GGA ATG GGC CTT     313
Arg Glu Thr Asp Pro Lys Lys Arg Val Val Ile Thr Gly Met Gly Leu
            50                  55                  60

GTC TCC GTT TTC GGC TCC GAC GTC GAT GCG TAC TAC GAC AAG CTC CTG     361
Val Ser Val Phe Gly Ser Asp Val Asp Ala Tyr Tyr Asp Lys Leu Leu
        65                  70                  75

TCA GGC GAG AGC GGG ATC GGC CCA ATC GAC CGC TTC GAC GCC TCC AAG     409
Ser Gly Glu Ser Gly Ile Gly Pro Ile Asp Arg Phe Asp Ala Ser Lys
    80                  85                  90

TTC CCC ACC AGG TTC GGC GGC CAG ATT CGT GGC TTC AAC TCC ATG GGA     457
Phe Pro Thr Arg Phe Gly Gly Gln Ile Arg Gly Phe Asn Ser Met Gly
95                  100                 105                 110

TAC ATT GAC GGC AAA AAC GAC AGG CGG CTT GAT GAT TGC CTT CGC TAC     505
Tyr Ile Asp Gly Lys Asn Asp Arg Arg Leu Asp Asp Cys Leu Arg Tyr
                115                 120                 125
```

FIGURE 5
1/4

TGC ATT GTC GCC GGG AAG AAG TCT CTT GAG GAC GCC GAT CTC GGT GCC 553
Cys Ile Val Ala Gly Lys Lys Ser Leu Glu Asp Ala Asp Leu Gly Ala
130 135 140

GAC CGC CTC TCC AAG ATC GAC AAG GAG AGA GCC GGA GTG CTG GTT GGG 601
Asp Arg Leu Ser Lys Ile Asp Lys Glu Arg Ala Gly Val Leu Val Gly
145 150 155

ACA GGA ATG GGT GGT CTG ACT GTC TTC TCT GAC GGG GTT CAA TCT CTT 649
Thr Gly Met Gly Gly Leu Thr Val Phe Ser Asp Gly Val Gln Ser Leu
160 165 170

ATC GAG AAG GGT CAC CGG AAA ATC ACC CCT TTC TTC ATC CCC TAT GCC 697
Ile Glu Lys Gly His Arg Lys Ile Thr Pro Phe Phe Ile Pro Tyr Ala
175 180 185 190

ATT ACA AAC ATG GGG TCT GCC CTG CTC GCT ATT GAA CTC GGT CTG ATG 745
Ile Thr Asn Met Gly Ser Ala Leu Leu Ala Ile Glu Leu Gly Leu Met
195 200 205

GGC CCA AAC TAT TCA ATT TCC ACT GCA TGT GCC ACT TCC AAC TAC TGC 793
Gly Pro Asn Tyr Ser Ile Ser Thr Ala Cys Ala Thr Ser Asn Tyr Cys
210 215 220

TTC CAT GCT GCT GCT AAT CAT ATC CGC CGT GGT GAG GCT GAT CTT ATG 841
Phe His Ala Ala Ala Asn His Ile Arg Arg Gly Glu Ala Asp Leu Met
225 230 235

ATT GCT GGA GGC ACT GAG GCC GCA ATC ATT CCA ATT GGG TTG GGA GGC 889
Ile Ala Gly Gly Thr Glu Ala Ala Ile Ile Pro Ile Gly Leu Gly Gly
240 245 250

FIGURE 5
2/4

```
TTT GTG GCT TGC AGG GCT CTG TCT CAA AGG AAC GAT GAC CCT CAG ACT      937
Phe Val Ala Cys Arg Ala Leu Ser Gln Arg Asn Asp Asp Pro Gln Thr
255                 260                 265                 270

GCC TCT AGG CCC TGG GAT AAA GAC CGT GAT GGT TTT GTG ATG GGT GAA      985
Ala Ser Arg Pro Trp Asp Lys Asp Arg Asp Gly Phe Val Met Gly Glu
                275                 280                 285

GGT GCT GGA GTG TTG GTG CTG GAG AGC TTG GAA CAT GCA ATG AAA CGA     1033
Gly Ala Gly Val Leu Val Leu Glu Ser Leu Glu His Ala Met Lys Arg
            290                 295                 300

GGA GCA CCT ATT ATT GCA GAG TAT TTG GGA GGT GCA ATC AAC TGT GAT     1081
Gly Ala Pro Ile Ile Ala Glu Tyr Leu Gly Gly Ala Ile Asn Cys Asp
        305                 310                 315

GCT TAT CAC ATG ACT GAC CCA AGG GCT GAT GGT CTC GGT GTC TCC TCT     1129
Ala Tyr His Met Thr Asp Pro Arg Ala Asp Gly Leu Gly Val Ser Ser
    320                 325                 330

TGC ATT GAG AGT AGC CTT GAA GAT GCT GGC GTC TCA CCT GAA GAG GTC     1176
Cys Ile Glu Ser Ser Leu Glu Asp Ala Gly Val Ser Pro Glu Glu Val
335                 340                 345                 350

AAT TAC ATA AAT GCT CAT GCG ACT TCT ACT CTA GCT GGG GAT CTC GCC     1224
Asn Tyr Ile Asn Ala His Ala Thr Ser Thr Leu Ala Gly Asp Leu Ala
                355                 360                 365

GAG ATA AAT GCC ATC AAG AAG GTT TTC AAG AAC ACA AAG GAT ATC AAA     1272
Glu Ile Asn Ala Ile Lys Lys Val Phe Lys Asn Thr Lys Asp Ile Lys
            370                 375                 380
```

FIGURE 5
3/4

```
ATT AAT GCA ACT AAG TCA ATG ATC GGA CAC TGT CTT GGA GCC TCT GGA    1320
Ile Asn Ala Thr Lys Ser Met Ile Gly His Cys Leu Gly Ala Ser Gly
        385                 390                 395

GGT CTT GAA GCT ATA GCG ACT ATT AAG GGA ATA AAC ACC GGC TGG CTT    1368
Gly Leu Glu Ala Ile Ala Thr Ile Lys Gly Ile Asn Thr Gly Trp Leu
    400                 405                 410

CAT CCC AGC ATT AAT CAA TTC AAT CCT GAG CCA TCC GTG GAG TTC GAC    1416
His Pro Ser Ile Asn Gln Phe Asn Pro Glu Pro Ser Val Glu Phe Asp
415                 420                 425                 430

ACT GTT GCC AAC AAG AAG CAG CAA CAC GAA GTT AAT GTT GCG ATC TCG    1464
Thr Val Ala Asn Lys Lys Gln Gln His Glu Val Asn Val Ala Ile Ser
                435                 440                 445

AAT TCA TTT GGA TTC GGA GGC CAC AAC TCA GTC GTG GCT TTC TCG GCT    1512
Asn Ser Phe Gly Phe Gly Gly His Asn Ser Val Val Ala Phe Ser Ala
            450                 455                 460

TTC AAG CCA TGA TTACC CATTTCACAA GGCACTTGTC ATTGAGAGTA CGGTTGTTCG  1569
Phe Lys Pro
        465

TCAAACCCAT TTAGGATACT GTTCTATGTA AAAAAAAGTA AGGATTATCA CTTTCCCTTC  1629

TAATCCTGTC TCCAGTTTGA GAATGAAATT ATATTTATTT TAAAAAAAAA AAAAAAGGGC  1689

GGCCGCTCTA GAGGATCCAA GCT                                          1712
```

FIGURE 5
4/4

Sequence Range: 1 to 1802

```
        10         20         30         40         50         60
                                                                 *
GGTCGACCCA CGCGTCCGGG CTTTCCGACC ACATTTCATT TCTTGCCTCG TTATCTCCGC

        70         80         90           100         110
CGCTCCTCCG CCGTCGTTCG CCGCCGCCGC C ATG CAA TCC CTC CAC TCC CCT TCC
                                   Met Gln Ser Leu His Ser Pro Ser

  120          130          140         150         160
   *
CTC CGC CCC TCC CCT CTC GAG CCC TTC CGC CTC AAT TCC CCC TCC TCC
Leu Arg Pro Ser Pro Leu Glu Pro Phe Arg Leu Asn Ser Pro Ser Ser

     170         180         190          200         210
                   *
GCC GCC GCT CTC CGC CCC CTC CGT CGC GCC AGC CTC CCC GTC ATC CGT
Ala Ala Ala Leu Arg Pro Leu Arg Arg Ala Ser Leu Pro Val Ile Arg

       220         230         240         250
                                 *
GCT GCC ACC GCC TCC GCC CCC AAG CGC GAG TCC GAC CCC AAG AAG CGG
Ala Ala Thr Ala Ser Ala Pro Lys Arg Glu Ser Asp Pro Lys Lys Arg

260         270         280         290         300
                                                  *
GTC GTC ATC ACC GGC ATG GGC CTC GTC TCC GTC TTC GGC TCC GAC GTC
Val Val Ile Thr Gly Met Gly Leu Val Ser Val Phe Gly Ser Asp Val

310         320         330         340         350
GAC GCC TAC TAC GAC AAG CTG CTC TCC GGC GAG AGC GGC ATC AGC CTA
Asp Ala Tyr Tyr Asp Lys Leu Leu Ser Gly Glu Ser Gly Ile Ser Leu
```

FIGURE 6
1/5

```
   360        370        380        390        400
     *
ATC GAC CGC TTC GAC GCT TCC AAA TTC CCC ACC AGG TTC GCC GGC CAG
Ile Asp Arg Phe Asp Ala Ser Lys Phe Pro Thr Arg Phe Ala Gly Gln

       410        420        430        440        450
                    *
ATC CGT GGC TTC AAC GCG ACG GGC TAC ATC GAC GGC AAG AAC GAC CGG
Ile Arg Gly Phe Asn Ala Thr Gly Tyr Ile Asp Gly Lys Asn Asp Arg

           460        470        480        490
                                   *
CGG CTC GAC GAT TGC CTC CGC TAC TGC ATT GTC GCC GGC AAG AAG GCT
Arg Leu Asp Asp Cys Leu Arg Tyr Cys Ile Val Ala Gly Lys Lys Ala

500        510        520        530        540
                                             *
CTC GAA GAC GCC GAT CTC GCC GGC CAA TCC CTC TCC AAG ATT GAT AAG
Leu Glu Asp Ala Asp Leu Ala Gly Gln Ser Leu Ser Lys Ile Asp Lys

550        560        570        580        590
GAG AGG GCC GGA GTG CTA GTT GGA ACC GGT ATG GGT GGC CTA ACT GTC
Glu Arg Ala Gly Val Leu Val Gly Thr Gly Met Gly Gly Leu Thr Val

    600        610        620        630        640
     *
TTC TCT GAC GGG GTT CAG AAT CTC ATC GAG AAA GGT CAC CGG AAG ATC
Phe Ser Asp Gly Val Gln Asn Leu Ile Glu Lys Gly His Arg Lys Ile

       650        660        670        680        690
                    *
TCC CCG TTT TTC ATT CCA TAT GCC ATT ACA AAC ATG GGG TCT GCG CTG
Ser Pro Phe Phe Ile Pro Tyr Ala Ile Thr Asn Met Gly Ser Ala Leu
```

FIGURE 6
2/5

```
              700                 710                 720                 730
                                                        *
CTT GCC ATC GAT TTG GGT CTG ATG GGC CCA AAC TAT TCG ATT TCA ACT
Leu Ala Ile Asp Leu Gly Leu Met Gly Pro Asn Tyr Ser Ile Ser Thr

740                 750                 760                 770                 780
                                                                                   *
GCA TGT GCT ACT TCC AAC TAC TGC TTT TAT GCT GCC GCC AAT CAT ATC
Ala Cys Ala Thr Ser Asn Tyr Cys Phe Tyr Ala Ala Ala Asn His Ile

790                 800                 810                 820                 830
CGC CGA GGT GAG GCT GAC CTG ATG ATT GCT GGA GGA ACT GAG GCT GCG
Arg Arg Gly Glu Ala Asp Leu Met Ile Ala Gly Gly Thr Glu Ala Ala

    840                 850                 860                 870                 880
     *
GTC ATT CCA ATT GGT TTA GGA GGA TTC GTT GCC TGC AGG GCT TTA TCT
Val Ile Pro Ile Gly Leu Gly Gly Phe Val Ala Cys Arg Ala Leu Ser

        890                 900                 910                 920                 930
                              *
CAA AGG AAT GAT GAT CCT CAG ACT GCC TCA AGG CCG TGG GAT AAG GAC
Gln Arg Asn Asp Asp Pro Gln Thr Ala Ser Arg Pro Trp Asp Lys Asp

           940                 950                 960                 970
                                                     *
CGT GAT GGC TTT GTG ATG GGT GAA GGG GCT GGA GTA TTG GTT ATG GAG
Arg Asp Gly Phe Val Met Gly Glu Gly Ala Gly Val Leu Val Met Glu

980                 990                 1000                1010                1020
                                                                                   *
AGC TTG GAG CAT GCA ATG AAA CGG GGA GCG CCG ATT ATT GCA GAA TAT
Ser Leu Glu His Ala Met Lys Arg Gly Ala Pro Ile Ile Ala Glu Tyr
```

FIGURE 6
3/5

```
1030      1040      1050      1060      1070
TTG GGA GGT GCA GTC AAC TGT GAT GCT TAT CAT ATG ACT GAT CCA AGG
Leu Gly Gly Ala Val Asn Cys Asp Ala Tyr His Met Thr Asp Pro Arg

  1080      1090      1100      1110      1120
     *
GCT GAT GGG CTT GGT GTC TCC TCG TGC ATT GAG AGC AGT CTC GAA GAT
Ala Asp Gly Leu Gly Val Ser Ser Cys Ile Glu Ser Ser Leu Glu Asp

     1130      1140      1150      1160      1170
                  *
GCC GGG GTC TCA CCT GAA GAG GTC AAT TAC ATA AAT GCT CAT GCG ACT
Ala Gly Val Ser Pro Glu Glu Val Asn Tyr Ile Asn Ala His Ala Thr

      1180      1190      1200      1210
                             *
TCT ACT CTT GCT GGG GAT CTT GCC GAG ATA AAT GCC ATT AAG AAA GTT
Ser Thr Leu Ala Gly Asp Leu Ala Glu Ile Asn Ala Ile Lys Lys Val

1220      1230      1240      1250      1260
                                           *
TTC AAG AAC ACC AAG GAA ATC AAA ATC AAT GCA ACT AAG TCA ATG ATC
Phe Lys Asn Thr Lys Glu Ile Lys Ile Asn Ala Thr Lys Ser Met Ile

1270      1280      1290      1300      1310
GGA CAC TGT CTT GGA GCA TCA GGA GGT CTT GAA GCC ATC GCA ACC ATT
Gly His Cys Leu Gly Ala Ser Gly Gly Leu Glu Ala Ile Ala Thr Ile

  1320      1330      1340      1350      1360
     *
AAG GGA ATA ACC ACC GGC TGG CTT CAT CCC AGC ATT AAT CAA TTT AAT
Lys Gly Ile Thr Thr Gly Trp Leu His Pro Ser Ile Asn Gln Phe Asn
```

FIGURE 6
4/5

```
    1370        1380        1390        1400        1410
                  *
CCC GAG CCA TCG GTG GAC TTC AAC ACT GTT GCC AAC AAA AAG CAG CAA
Pro Glu Pro Ser Val Asp Phe Asn Thr Val Ala Asn Lys Lys Gln Gln

     1420        1430        1440        1450
                               *
CAT GAA GTG AAC GTC GCT ATC TCG AAT TCT TTT GGA TTT GGA GGG CAC
His Glu Val Asn Val Ala Ile Ser Asn Ser Phe Gly Phe Gly Gly His

1460        1470        1480        1490        1500        1510
                                                   *
AAC TCG GTT GTG GCA TTC TCA GCT TTC AAG CCA TGA ATTCT ACTTGGTTCA
Asn Ser Val Val Ala Phe Ser Ala Phe Lys Pro ***

      1520       1530       1540       1550       1560       1570
                                                     *
AAATGCACAC CAGTTGCTGA GATAGGGCTT CAACTTGCAG AGCAATTTTT TAAATGCCTT

      1580       1590       1600       1610       1620       1630
                                                     *
GTCGGAAGAG CGTAATACCG GAATAGGTCG GTCCTTTGAT AGTTCCTCGA AGCCATTTAG

      1640       1650       1660       1670       1680       1690
                                                     *
GATGATGTTT TACTGTAATA ATCGAAGATG ATTCCCATTT TAAATCTAGT CTCTGATTTA

      1700       1710       1720       1730       1740       1750
                                                     *
TGTATTAGAA AGACCAATGA AAGATTTTGT GTCATGTTTG TGTTGTCAAT GTTATTTAAG

      1760       1770       1780       1790       1800
                                                     *
ATAAAGCAAA AAAAAAAAAA AAGGGCGGCC GCTCTAGAGG ATCCAGCTTA CT
```

FIGURE 6

5/5

```
Sequence Range: 1 to 2369

        10         20         30         40         50         60
                                                                  *
GTACGCCTGC AGGTACCGGT CCGGAATTCC CGGGTCGACC CACGCGTCCG CATAAAAGAG

        70         80         90        100        110        120
                                                                  *
AGAGAGAGGG ATCCATCGAA TGCGGCCACC CTCCTTTCAT CTTCGATTCA TTACCATACC

       130        140        150        160        170        180
                                                                  *
ATTCCGCTGA TCCATTTTCC GCCTTTTCCG GGTCTTTCAT CCCAAAGGGT ATCCTTTTCT

       190        200        210           220           230
ATCCTATCTT CTCAAAGGGT CAGTCAGTTC CCTCCA ATG CCT GCC GCC TCT TCC
                                        Met Pro Ala Ala Ser Ser>

   240             250              260              270              280
    *
CTG CTC GCT TCC CCT CTC TGT ACG TGG CTC CTT GCC GCC TGC ATG TCT
Leu Leu Ala Ser Pro Leu Cys Thr Trp Leu Leu Ala Ala Cys Met Ser>

       290             300              310              320              330
                        *
ACC TCC TTC CAC CCC TCC GAC CCT CTT CCG CCT TCC ATC TCC TCT CCT
Thr Ser Phe His Pro Ser Asp Pro Leu Pro Pro Ser Ile Ser Ser Pro>

           340              350              360              370
                                              *
CGC CGA CGC CTC TCC CGC CGC CGG ATT CTC TCC CAA TGC GCC CCA CTA
Arg Arg Arg Leu Ser Arg Arg Arg Ile Leu Ser Gln Cys Ala Pro Leu>
```

FIGURE 7
1/7

32/66

```
380          390          400          410          420
                                                     *
CCT TCT GCT TCC TCC GCC CTC CGC GGA TCC AGT TTC CAT ACC CTC GTC
Pro Ser Ala Ser Ser Ala Leu Arg Gly Ser Ser Phe His Thr Leu Val>

  430          440          450          460          470
ACC TCT TAC CTC GCC TGC TTC GAG CCC TGC CAT GAC TAC TAT ACA TCC
Thr Ser Tyr Leu Ala Cys Phe Glu Pro Cys His Asp Tyr Tyr Thr Ser>

    480          490          500          510          520
     *
GCA TCC TTG TTC GGA TCC AGA CCC ATT CGC ACC ACC CGC AGG CAC CGG
Ala Ser Leu Phe Gly Ser Arg Pro Ile Arg Thr Thr Arg Arg His Arg>

       530          540          550          560          570
                     *
AGG CTC AAT CGA GCT TCC CCT TCC AGG GAG GCA ATG GCC GTG GCT CTG
Arg Leu Asn Arg Ala Ser Pro Ser Arg Glu Ala Met Ala Val Ala Leu>

          580          590          600          610
                                     *
CAA CCT GAA CAG GAA GTT ACC ACA AAG AAG AAG CCA AGT ATC AAA CAG
Gln Pro Glu Gln Glu Val Thr Thr Lys Lys Lys Pro Ser Ile Lys Gln>

620          630          640          650          660
                                                     *
CGG CGA GTA GTT GTG ACT GGA ATG GGT GTG GTG ACT CCT CTA GGC CAT
Arg Arg Val Val Val Thr Gly Met Gly Val Val Thr Pro Leu Gly His>

  670          680          690          700          710
GAC CCT GAT GTT TTC TAC AAT AAT CTG CTT GAT GGA ACG AGT GGC ATA
Asp Pro Asp Val Phe Tyr Asn Asn Leu Leu Asp Gly Thr Ser Gly Ile>
```

FIGURE 7
2/7

```
        720                 730                 740                 750                 760
          *
AGC GAG ATA GAG ACC TTT GAT TGT GCT CAA TTT CCT ACG AGA ATT GCT
Ser Glu Ile Glu Thr Phe Asp Cys Ala Gln Phe Pro Thr Arg Ile Ala>

            770                 780                 790                 800                 810
                                  *
GGA GAG ATC AAG TCT TTC TCC ACA GAT GGT TGG GTG GCC CCG AAG CTC
Gly Glu Ile Lys Ser Phe Ser Thr Asp Gly Trp Val Ala Pro Lys Leu>

                820                 830                 840                 850
                                                          *
TCT AAG AGG ATG GAC AAG TTC ATG CTA TAC ATG CTG ACC GCT GGC AAG
Ser Lys Arg Met Asp Lys Phe Met Leu Tyr Met Leu Thr Ala Gly Lys>

860                 870                 880                 890                 900
                                                                                  *
AAA GCA TTA ACA GAT GGT GGA ATC ACC GAA GAT GTG ATG AAA GAG CTA
Lys Ala Leu Thr Asp Gly Gly Ile Thr Glu Asp Val Met Lys Glu Leu>

    910                 920                 930                 940                 950
GAT AAA AGA AAA TGC GGA GTT CTC ATT GGC TCA GCA ATG GGT GGA ATG
Asp Lys Arg Lys Cys Gly Val Leu Ile Gly Ser Ala Met Gly Gly Met>

        960                 970                 980                 990                 1000
          *
AAG GTA TTC AAT GAT GCC ATT GAA GCC CTA AGG ATT TCA TAT AAG AAG
Lys Val Phe Asn Asp Ala Ile Glu Ala Leu Arg Ile Ser Tyr Lys Lys>

          1010                1020                1030                1040                1050
                                *
ATG AAT CCC TTT TGT GTA CCT TTC GCT ACC ACA AAT ATG GGA TCA GCT
Met Asn Pro Phe Cys Val Pro Phe Ala Thr Thr Asn Met Gly Ser Ala>
```

FIGURE 7
3/7

```
           1060        1070        1080        1090
                                     *
ATG CTT GCA ATG GAC TTG GGA TGG ATG GGG CCC AAC TAC TCG ATA TCT
Met Leu Ala Met Asp Leu Gly Trp Met Gly Pro Asn Tyr Ser Ile Ser>

1100        1110        1120        1130        1140
                                                  *
ACT GCT TGT GCA ACG AGT AAC TTT TGT ATA ATG AAT GCT GCG AAC CAT
Thr Ala Cys Ala Thr Ser Asn Phe Cys Ile Met Asn Ala Ala Asn His>

 1150        1160        1170        1180        1190
ATA ATC AGA GGC GAA GCA GAT GTG ATG CTT TGC GGG GGC TCA GAT GCG
Ile Ile Arg Gly Glu Ala Asp Val Met Leu Cys Gly Gly Ser Asp Ala>

  1200        1210        1220        1230        1240
   *
GTA ATC ATA CCT ATT GGT ATG GGA GGT TTT GTT GCA TGC CGA GCT TTG
Val Ile Ile Pro Ile Gly Met Gly Gly Phe Val Ala Cys Arg Ala Leu>

    1250        1260        1270        1280        1290
                  *
TCC CAG AGA AAT TCC GAC CCT ACT AAA GCT TCA AGA CCA TGG GAC AGT
Ser Gln Arg Asn Ser Asp Pro Thr Lys Ala Ser Arg Pro Trp Asp Ser>

      1300        1310        1320        1330
                                *
AAT CGT GAT GGA TTT GTT ATG GGG GAA GGA GCT GGA GTG CTA CTA CTA
Asn Arg Asp Gly Phe Val Met Gly Glu Gly Ala Gly Val Leu Leu Leu>

1340        1350        1360        1370        1380
                                                  *
GAG GAG TTG GAG CAT GCA AAG AAA AGA GGT GCG ACT ATT TAC GCA GAA
Glu Glu Leu Glu His Ala Lys Lys Arg Gly Ala Thr Ile Tyr Ala Glu>
```

FIGURE 7
4/7

```
1390       1400       1410       1420       1430
TTT CTA GGT GGG AGT TTC ACT TGC GAT GCC TAC CAC ATG ACC GAG CCT
Phe Leu Gly Gly Ser Phe Thr Cys Asp Ala Tyr His Met Thr Glu Pro>

  1440       1450       1460       1470       1480
     *
CAC CCT GAT GGA GCT GGA GTG ATT CTC TGC ATA GAG AAG GCT TTG GCT
His Pro Asp Gly Ala Gly Val Ile Leu Cys Ile Glu Lys Ala Leu Ala>

     1490       1500       1510       1520       1530
                   *
CAG TCA GGA GTC TCT AGG GAA GAC GTA AAT TAC ATA AAT GCC CAT GCC
Gln Ser Gly Val Ser Arg Glu Asp Val Asn Tyr Ile Asn Ala His Ala>

        1540       1550       1560       1570
                                 *
ACA TCC ACT CCG GCT GGA GAT ATC AAA GAG TAC CAA GCT CTT ATC CAC
Thr Ser Thr Pro Ala Gly Asp Ile Lys Glu Tyr Gln Ala Leu Ile His>

1580       1590       1600       1610       1620
                                             *
TGT TTC GGC CAA AAC AGA GAG TTA AAA GTT AAT TCA ACC AAA TCA ATG
Cys Phe Gly Gln Asn Arg Glu Leu Lys Val Asn Ser Thr Lys Ser Met>

 1630       1640       1650       1660       1670
ATT GGT CAC CTT CTC GGA GCA GCC GGT GGT GTG GAA GCA GTT TCA GTA
Ile Gly His Leu Leu Gly Ala Ala Gly Gly Val Glu Ala Val Ser Val>

   1680       1690       1700       1710       1720
      *
GTT CAG GCA ATA AGG ACT GGG TGG ATC CAT CCG AAT ATT AAT TTG GAA
Val Gln Ala Ile Arg Thr Gly Trp Ile His Pro Asn Ile Asn Leu Glu>
```

FIGURE 7
5/7

36166

```
      1730          1740          1750          1760          1770
                       *
AAC CCA GAT GAA GGC GTG GAT ACA AAA TTG CTC GTG GGT CCT AAG AAG
Asn Pro Asp Glu Gly Val Asp Thr Lys Leu Leu Val Gly Pro Lys Lys>

        1780          1790          1800          1810
                                       *
GAG AGA CTG AAC GTT AAG GTC GGT TTG TCT AAT TCA TTT GGG TTT GGT
Glu Arg Leu Asn Val Lys Val Gly Leu Ser Asn Ser Phe Gly Phe Gly>

1820          1830          1840          1850          1860          1870
                                                           *
GGG CAC AAC TCG TCC ATA CTC TTC GCC CCT TAC ATC TAG GAC GTTTCCGTGT
Gly His Asn Ser Ser Ile Leu Phe Ala Pro Tyr Ile ***>

      1880        1890        1900        1910        1920        1930
                                                         *
GTGGAATTCT ACTCAACATA TCAAAGCTGA AGTTTTGAGG ACTCCAGCAT GTTGGTAGCT

      1940        1950        1960        1970        1980        1990
                                                         *
CCTTACGTCT CTAGACATGC CCATGAGTTT TGTGTCCGGA GCTTTAGTCG GAACCATGAC

      2000        2010        2020        2030        2040        2050
                                                         *
GGATTGAGTA CTCATGGCGA CACTTGATAT ACTCCTTGCT AGAATTGTTG GTAGAGCAAT

      2060        2070        2080        2090        2100        2110
                                                         *
ATTCATTATC TCATATTTTT TTTTTCTCTG AAATCTCCCT CCTTGCAATA GTTGTACTTT

      2120        2130        2140        2150        2160        2170
                                                         *
CGAGCTTTTC ATCGAGTCAG TGAAGAAGAG AACAAAGCTG TTAACTCGGG CACGTAGTAA
```

FIGURE 7
6/7

```
      2180       2190       2200       2210       2220       2230
                                                     *
CCATTTGCCC TTTGTTTTGC TCTCTATTTC ATCACCGTTT TGTGGTTTTA AAATTTGTAA

      2240       2250       2260       2270       2280       2290
                                                     *
AACTAGAAGA CTGGTTTAGA TTGGTTTGTT TTCTCATTGA TAATTGGGGR ATGTATGTTT

      2300       2310       2320       2330       2340       2350
                                                     *
TGGAAATAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA

      2360
AGGGCGGCCG CTCTAGAGG
```

FIGURE 7
7/7

Sequence Range: 1 to 2374

```
        10         20         30         40         50         60
                                                                  *
-A-CNTGGTC CGGAATTCCC GGGTCGACCC ACGCGTCCGC GACGCCAACC CACACCAAAC

        70         80         90        100        110        120
                                                                  *
TTCCTCAGCT TCTCTTCTCA AGACGGACGC CATTGGCAGC AGACAGACAG ACAGACAGAC

       130        140        150        160        170        180
                                                                  *
CCATAAAAGA GAGAGAGAGG GATCCATCGA ATGCGGCCAC CCTCCTTTCA TCTTCGATTC

       190        200        210        220        230        240
                                                                  *
ATTACCATAC CATTCCGCTG ATCCATTTTC CGCCTTTTCC GGGTCTTTCA TCCCAAAGGG

       250        260        270        280        290        300
                                                                  *
TATCCTTTTC TATCCTATCT TCTCAAAGGG TCAGTCAGTT CCCTCCAATG CCTGCCGCCT

       310        320        330        340        350        360
                                                                  *
CTTCCCTGCT CGCTTCCCCT CTCTGTACGT GGCTCCTTGC CGCCTGCATG TCTACCTCCT

       370        380        390        400        410        420
                                                                  *
TCCACCCCTC CGACCCTCTT CCGCCTTCCA TCTCCTCTCC TCGCCGACGC CTCTCCCGCC

       430        440        450        460        470        480
                                                                  *
GCCGGATTCT CTCCCAATGC GCCCCACTAC CTTCTGCTTC CTCCGCCCTC CGCGGATCCA
```

FIGURE 8
1/5

39/166

```
       490        500        510        520        530        540
                                                                  *
GTTTCCATAC CCTCGTCACC TCTTACCTCG CCTGCTTCGA GCCCTGCCAT GACTACTATA

       550        560        570        580        590        600
                                                                  *
CATCCGCATC CTTGTTCGGA TCCAGACCCA TTCGCACCAC CCGCAGGCAC CGGAGGCTCA

       610        620        630        640        650        660
                                                                  *
ATCGAGCTTC CCCTTCCAGG GGAGGCAATG GCCGTGGCTC TGCAACCTGA ACAGGAAGTT

       670        680        690        700        710        720
                                                                  *
ACCACAAAGA AGAAGCCAAG TATCAAACAG CGGCGAGTAG TTGTGACTGG AATGGGTGTG

       730        740        750        760        770        780
                                                                  *
GTGACTCCTC TAGGCCATGA ACCTGATGTT TTTCTACAAT AATCTGCTTG ATGGAACGAG

       790        800        810        820        830        840
                                                                  *
TGGCATAAGC GAGATAGAGA CCTTTGATTG TGCTCAATTT CCTACGAGAA TTGCTGGAGA

       850        860        870        880        890        900
                                                                  *
GATCAAGTCT TTCTCCACAG ATGGTTGGGT GGCCCCGAAG CTCTCTAAGA GGATGGACAA

       910        920        930        940        950        960
                                                                  *
GTTCATGCTA TACATGCTGA CTGCTGGCAA GAAAGCATTA ACAGATGGTG GAATCACCGA

       970        980        990       1000       1010       1020
                                                                  *
AGATGTGATG AAAGAGCTAG ATAAAAGAAA ATGCGGAGTT CTCATTGGCT CAGCAATGGG
```

FIGURE 8
2/5

```
      1030       1040       1050       1060       1070       1080
                                                                *
TGGAATGAAG GTATTCAATG ATGCCATTGA AGCCCTAAGG ATTTCATATA AGAAGATGAA

      1090       1100       1110       1120       1130       1140
                                                                *
TCCCTTTTGT GTACCTTTCG CTACCACAAA TATGGGATCA GCTATGCTTG CAATGGACTT

      1150       1160       1170       1180       1190       1200
                                                                *
GGGATGGATG GGGCCCAACT ACTCGATATC TACTGCTTGT GCAACGAGTA ACTTTTGTAT

      1210       1220       1230       1240       1250       1260
                                                                *
AATGAATGCT GCGAACCATA TAATCAGAGG CGAAGCAGAT GTGATGCTTT GCGGGGGCTC

      1270       1280       1290       1300       1310       1320
                                                                *
AGATGCGGTA ATCATACCTA TTGGTATGGG AGGTTTTGTT GCATGCCGAG CTTTGTCCCA

      1330       1340       1350       1360       1370       1380
                                                                *
GAGAAATTCC GACCCTACTA AAGCTTCAAG ACCATGGGAC AGTAATCGTG ATGGATTTGT

      1390       1400       1410       1420       1430       1440
                                                                *
TATGGGGGAA GGAGCTGGAG TGCTACTACT AGAGGAGTTG GAGCATGCAA AGAAAAGAGG

      1450       1460       1470       1480       1490       1500
                                                                *
TGCGACTATT TACGCAGAAT TTCTAGGTGG GAGTTTCACT TGCGATGCCT ACCACATGAC
```

FIGURE 8
3/5

```
      1510       1520       1530       1540       1550       1560
                                                                *
CGAGCCTCAC CCTGATGGAG CTGGAGTGAT TCTCTGCATA GAGAAGGCTT TGGCTCAGTC

      1570       1580       1590       1600       1610       1620
                                                                *
AGGAGTCTCT AGGGAAGACG TAAATTACAT AAATGCCCAT GCCACATCCA CTCCGGCTGG

      1630       1640       1650       1660       1670       1680
                                                                *
AGATATCAAA GAGTACCAAG CTCTTATCCA CTGTTTCGGC CAAAACAGAG AGTTAAAAGT

      1690       1700       1710       1720       1730       1740
                                                                *
TAATTCAACC AAATCAATGA TTGGTCACCT TCTCGGAGCA GCCGGTGGTG TGGAAGCAGT

      1750       1760       1770       1780       1790       1800
                                                                *
TTCAGTAGTT CAGGCAATAA GGACTGGGTG GATCCATCCG AATATTAATT TGGAAAACCC

      1810       1820       1830       1840       1850       1860
                                                                *
AGATGAAGGC GTGGATACAA AATTGCTCGT GGGTCCTAAG AAGGAGAGAC TGAACGTTAA

      1870       1880       1890       1900       1910       1920
                                                                *
GGTCGGTTTG TCTAATTCAT TTGGGTTTGG TGGGCACAAC TCGTCCATAC TCTTCGCCCC

      1930       1940       1950       1960       1970       1980
                                                                *
TTACATCTAG GACGTTTCGT GTGTGGAATT CTACTCAACA TATCAAAGCT GAAGTTTTGA

      1990       2000       2010       2020       2030       2040
                                                                *
GGACTCCAGC ATGTTGGTAG CTCCTTACGT CTCTAGACAT GCCCATGAGT TTTGTGTCCG
```

FIGURE 8
4/5

42/66

```
      2050       2060       2070       2080       2090       2100
                                                                *
GAGCTTTAGT CGGAACCATG ACGGATTGAG TACTCATGGC GACACTTGAT ATACTCCTTG

      2110       2120       2130       2140       2150       2160
                                                                *
CTAGAATTGT TGGTAGAGCA ATATTCATTA TCTCATATTT TTTTTTTCTC TGAAATCTCC

      2170       2180       2190       2200       2210       2220
                                                                *
CTCCTTGCAA TAGTTGTACT TTCGAGCTTT TCATCGAGTC AGTGAAGAAG AGAACAAAGC

      2230       2240       2250       2260       2270       2280
                                                                *
TGTTAACTCG GGCACGTAGT AACCATTTGC CCTTTGTTTT GCTCTCTATT TCATCACCGT

      2290       2300       2310       2320       2330       2340
                                                                *
TTTGTGGTTT TAAAATTTGT AAAACTAGAA GACTGGTTTA GATTGGTTTG TTTTCTCAAA

      2350       2360       2370
AAAAAAAAAA AAGGGCGGCC GCTCTAGAGG ATCC
```

FIGURE 8
5/5

Sequence Range: 1 to 1580

```
        10         20         30          40         50
CCTGAATCGG ATTCAAGAGA GAGTTTCGTT GCTGGG ATG GCG AAT GCA TCT GGG
                                        Met Ala Asn Ala Ser Gly>

    60         70         80         90        100
     *
TTT CTG GGT TCT TCA GTT CCT GCC CTG AGA AGG GCA ACT CAG CAT TCG
Phe Leu Gly Ser Ser Val Pro Ala Leu Arg Arg Ala Thr Gln His Ser>

      110        120        130        140        150
                   *
ATT TCA TCG TCT CGT GGA TCT TCC TCG GAG TTT GTC TCC AAA AGG GTG
Ile Ser Ser Ser Arg Gly Ser Ser Ser Glu Phe Val Ser Lys Arg Val>

         160        170        180        190
                                 *
TTT TGC TGT AGT GCC GTT CAG GAT TCT GAC AGG CAG TCT TTG GGT GAT
Phe Cys Cys Ser Ala Val Gln Asp Ser Asp Arg Gln Ser Leu Gly Asp>

200        210        220        230        240
                                              *
TCT CGC TCG CCG AGG CTT GTG AGT AGA GGA TGC AAA TTA ATT GGA TCT
Ser Arg Ser Pro Arg Leu Val Ser Arg Gly Cys Lys Leu Ile Gly Ser>

 250        260        270        280        290
GGT TCT GCT ATA CCA GCT CTT CAA GTC TCA AAT GAT GAT CTT GCT AAA
Gly Ser Ala Ile Pro Ala Leu Gln Val Ser Asn Asp Asp Leu Ala Lys>

    300        310        320        330        340
     *
ATT GTC GAC ACC AAT GAT GAA TGG ATT ACT GTC CGA ACG GGG ATC CGC
Ile Val Asp Thr Asn Asp Glu Trp Ile Thr Val Arg Thr Gly Ile Arg>
```

FIGURE 9
1/5

```
          350                 360                 370                 380                 390
                               *
AAC CGA AGG GTT CTC TCA GGT AAA GAT AGT CTT ACA AAT TTA GCA TCA
Asn Arg Arg Val Leu Ser Gly Lys Asp Ser Leu Thr Asn Leu Ala Ser>

              400                 410                 420                 430
                                                       *
GAG GCA GCA AGG AAA GCT CTA GAG ATG GCA CAG GTA GAC GCA AAT GAT
Glu Ala Ala Arg Lys Ala Leu Glu Met Ala Gln Val Asp Ala Asn Asp>

440                 450                 460                 470                 480
                                                                                 *
GTG GAT ATG GTT TTG ATG TGT ACT TCT ACC CCT GAG GAC CTT TTC GGC
Val Asp Met Val Leu Met Cys Thr Ser Thr Pro Glu Asp Leu Phe Gly>

   490                 500                 510                 520                 530
AGT GCT CCT CAG ATA TCG AAA GCA CTT GGC TGC AAA AAG AAT CCT TTG
Ser Ala Pro Gln Ile Ser Lys Ala Leu Gly Cys Lys Lys Asn Pro Leu>

      540                 550                 560                 570                 580
       *
TCT TAC GAC ATT ACC GCT GCA TGC AGT GGA TTT GTG TTG GGT TTA GTC
Ser Tyr Asp Ile Thr Ala Ala Cys Ser Gly Phe Val Leu Gly Leu Val>

          590                 600                 610                 620                 630
                               *
TCA GCT GCT TGC CAC ATT AGA GGT GGG GGT TTT AAC AAT ATT CTA GTG
Ser Ala Ala Cys His Ile Arg Gly Gly Gly Phe Asn Asn Ile Leu Val>

              640                 650                 660                 670
                                                       *
ATT GGT GCT GAT TCT CTT TCT CGG TAT GTT GAC TGG ACC GAT CGG GGA
Ile Gly Ala Asp Ser Leu Ser Arg Tyr Val Asp Trp Thr Asp Arg Gly>
```

FIGURE 9

2/5

```
680       690       700       710       720
                                          *
ACA TGT ATT CTC TTT GGA GAT GCT GCT GGA GCT GTA GTG GTG CAG TCA
Thr Cys Ile Leu Phe Gly Asp Ala Ala Gly Ala Val Val Val Gln Ser>

  730       740       750       760       770
TGT GAT GCT GAG GAA GAT GGG CTC TTT GCT TTT GAT TTG CAT AGC GAT
Cys Asp Ala Glu Glu Asp Gly Leu Phe Ala Phe Asp Leu His Ser Asp>

    780       790       800       810       820
     *
GGA GAT GGG CAA AGG CAT CTA AAA GCT GCA ATC AAA GAA GAT GAA GTT
Gly Asp Gly Gln Arg His Leu Lys Ala Ala Ile Lys Glu Asp Glu Val>

       830       840       850       860       870
                  *
GAT AAA GCC CTG GGA CAT AAT GGG TCC ATC AGA GAT TTT CCA CCA AGG
Asp Lys Ala Leu Gly His Asn Gly Ser Ile Arg Asp Phe Pro Pro Arg>

          880       890       900       910
                                *
CGT TCT TCA TAC TCT TGC ATC CAA ATG AAC GGT AAA GAG GTA TTC CGC
Arg Ser Ser Tyr Ser Cys Ile Gln Met Asn Gly Lys Glu Val Phe Arg>

920       930       940       950       960
                                          *
TTT GCT TGC CGC TCT GTG CCT CAG TCA ATC GAA TCA GCA CTT GGA AAG
Phe Ala Cys Arg Ser Val Pro Gln Ser Ile Glu Ser Ala Leu Gly Lys>

  970       980       990       1000      1010
GCC GGT CTT AAT GGA TCC AAC ATC GAC TGG TTG CTG CTT CAT CAG GCA
Ala Gly Leu Asn Gly Ser Asn Ile Asp Trp Leu Leu Leu His Gln Ala>
```

FIGURE 9
3/5

```
  1020                1030               1040               1050               1060
     *
AAT CAG AGG ATC ATT GAT GCA GTA GCA ACA CGT CTA GAG GTT CCT CAA
Asn Gln Arg Ile Ile Asp Ala Val Ala Thr Arg Leu Glu Val Pro Gln>

        1070              1080               1090               1100               1110
                                 *
GAA CGA ATT ATC TCA AAC TTG GCA AAT TAC GGG AAC ACT AGT GCG GCA
Glu Arg Ile Ile Ser Asn Leu Ala Asn Tyr Gly Asn Thr Ser Ala Ala>

          1120               1130               1140               1150
                                                     *
TCC ATT CCC TTG GCA CTA GAC GAA GCT GTG AGG AGT GGA AAT GTG AAG
Ser Ile Pro Leu Ala Leu Asp Glu Ala Val Arg Ser Gly Asn Val Lys>

1160               1170               1180               1190               1200
                                                                                *
CCG GGT CAC GTG ATT GCA ACC GCA GGA TTT GGC GCC GGA CTC ACA TGG
Pro Gly His Val Ile Ala Thr Ala Gly Phe Gly Ala Gly Leu Thr Trp>

  1210               1220               1230               1240               1250               1260
                                                                                                      *
GGT TCT GCT ATT ATC AGG TGG GGA TAA GACTGAA GCCGAGCCAG CACTGCAGCT
Gly Ser Ala Ile Ile Arg Trp Gly ***>

      1270               1280               1290               1300               1310               1320
                                                                                                      *
TCCTCTCAAA CCGATGTTTC ACGAAATTTT GCTTCCATGA CCANAAAAAG AAGAAGTCAG

      1330               1340               1350               1360               1370               1380
                                                                                                      *
TCTTTTATGG AGCAAGCAAC ACGACACGAT CTTCATCACA TTGCCCTTTT TCGTTCCCCT
```

FIGURE 9

4/5

```
      1390       1400       1410       1420       1430       1440
                                                                 *
TTTCCATTAG TTTGATGATT TTGCTGACAA TACAATACCC ATAGTTTCTT TTGTCCCCAA

      1450       1460       1470       1480       1490       1500
                                                                 *
TAAGTTATTT GTTTCTTGTT TAATTGTTCA GCTTTTACTT CATTTTGTCT CGGGACATTG

      1510       1520       1530       1540       1550       1560
                                                                 *
GAGATGACAG CATAAACATC ATGTTTATAT TTTGCTAAAA AAAAAAAAAA AAAAAAAAAA

      1570       1580
AAAAAAAAAA AAAAAAAAAA
```

FIGURE 9
5/5

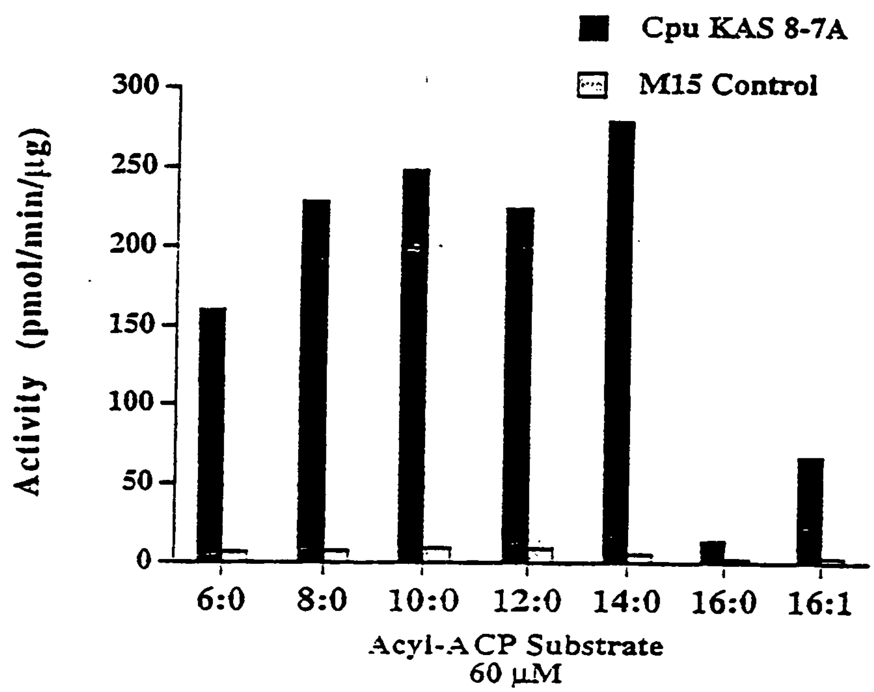

Cpu KAS 8-7A
M15 Control
300
250
200
150
100
50
0
Activity (pmol/min/µg)
6:0
8:0
10:0
12:0
14:0
16:0
16:1
Acyl-ACP Substrate
60 µM

FIGURE 10

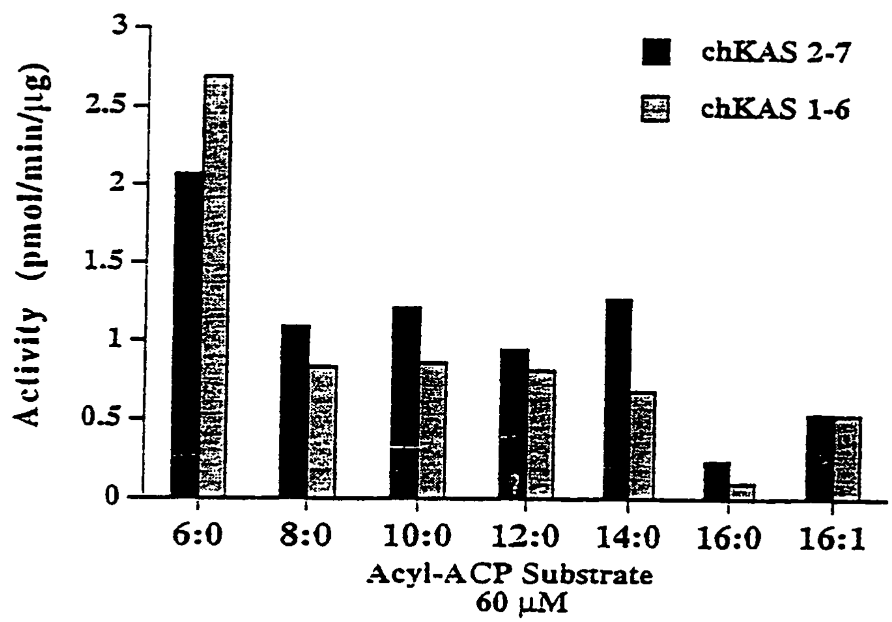

3
2.5
2
1.5
1
0.5
0
Activity (pmol/min/μg)
chKAS 2-7
chKAS 1-6
6:0
8:0
10:0
12:0
14:0
16:0
16:1
Acyl-ACP Substrate
60 μM

FIGURE 11

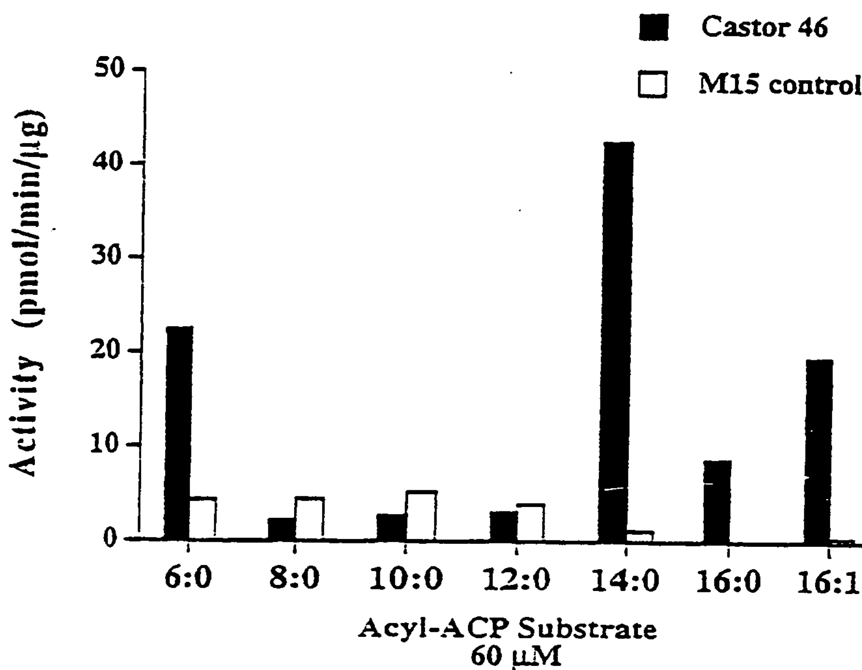
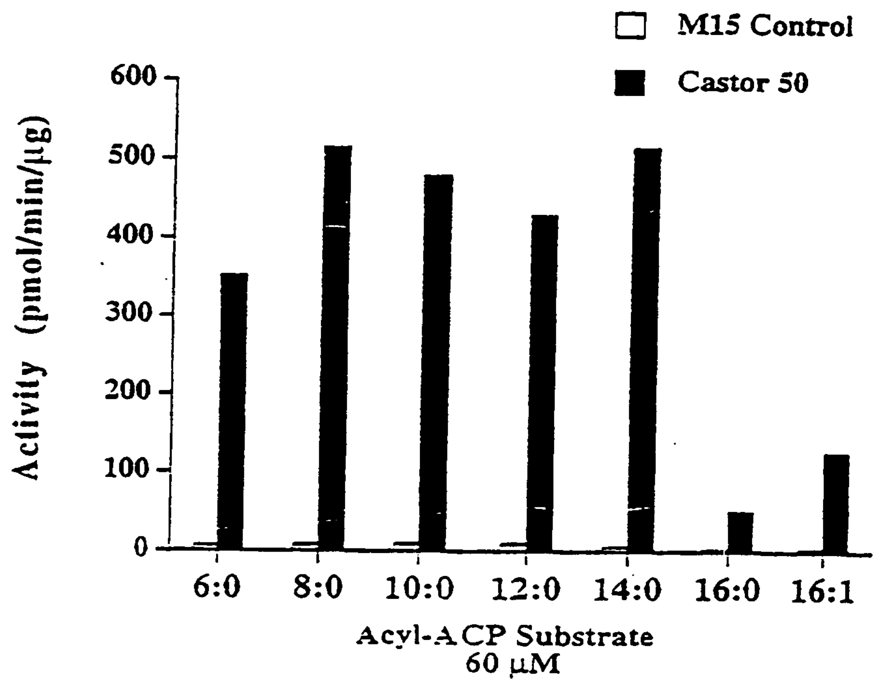

Castor 46
M15 control
50
40
30
20
10
0
Activity (pmol/min/µg)
6:0
8:0
10:0
12:0
14:0
16:0
16:1
Acyl-ACP Substrate
60 µM

FIGURE 12



E328013-28

FIGURE 13

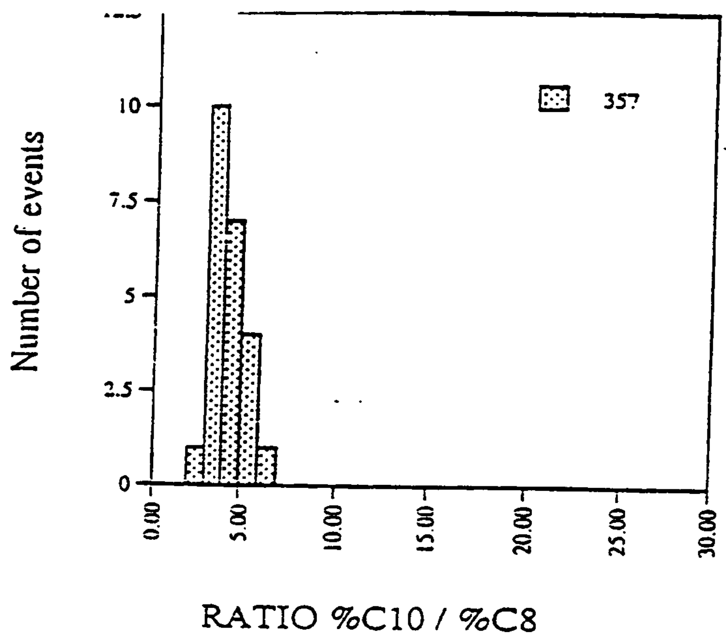

357
Number of events
10
7.5
5
2.5
0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
RATIO %C10 / %C8

FIGURE 15
1/2

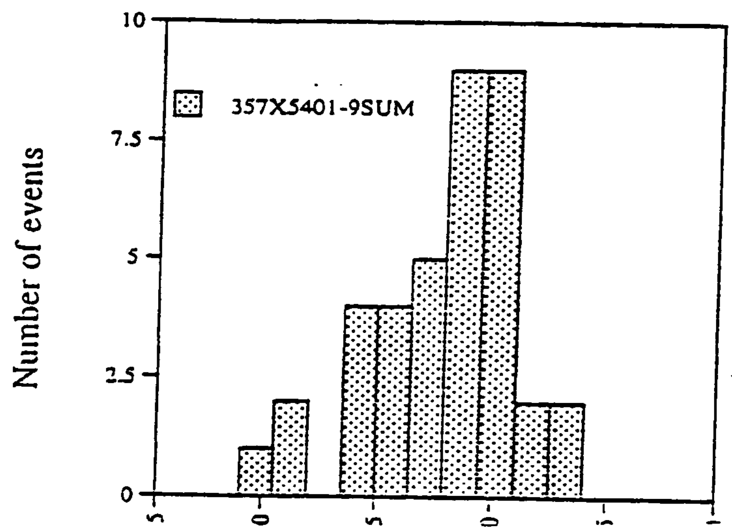


FIGURE 15
2/2

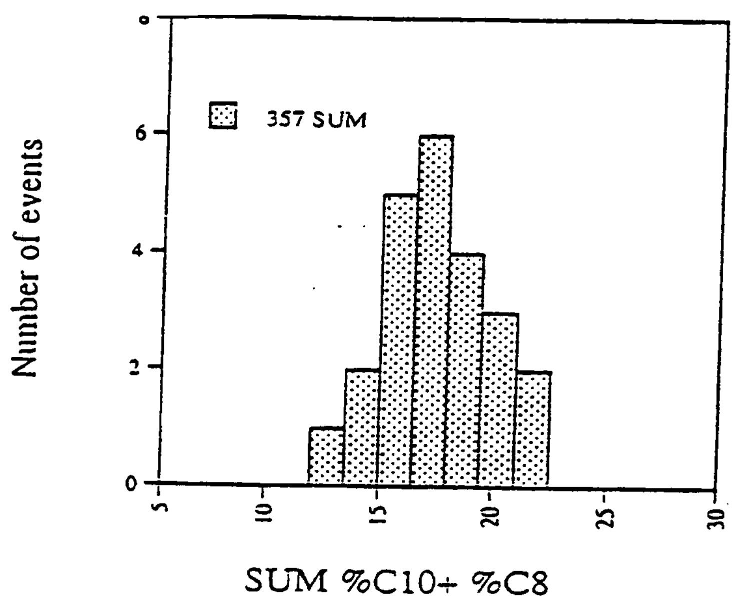
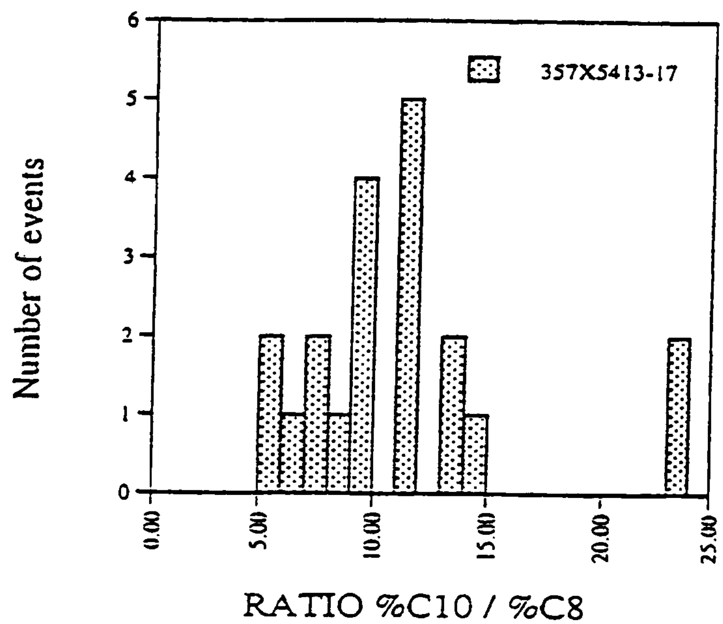


FIGURE 16



357X5413-17
Number of events
0
1
2
3
4
5
6
0.00
5.00
10.00
15.00
20.00
25.00
RATIO %C10 / %C8

FIGURE 17
1/2

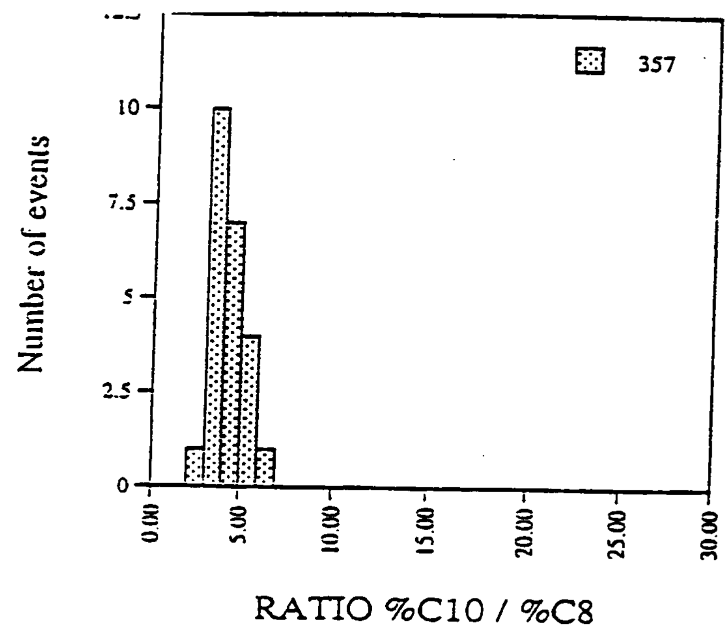

357
Number of events
10
7.5
5
2.5
0
0.00
5.00
10.00
15.00
20.00
25.00
30.00
RATIO %C10 / %C8

FIGURE 17

2/2

57/66

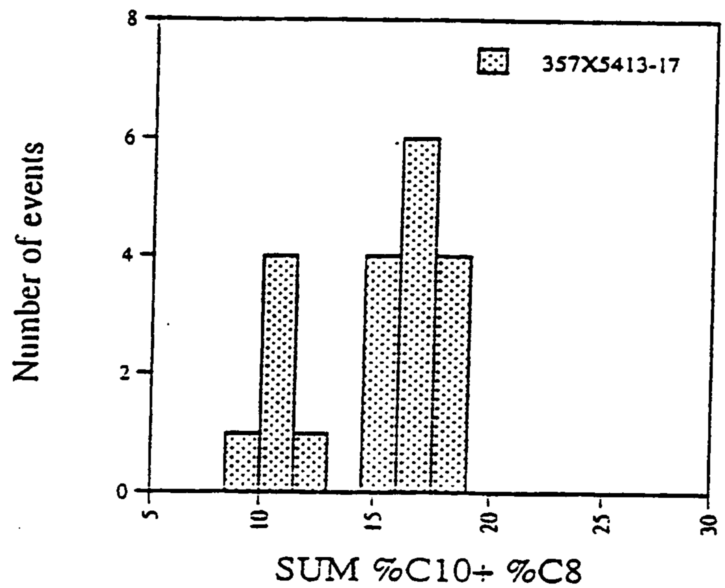


FIGURE 18
1/2

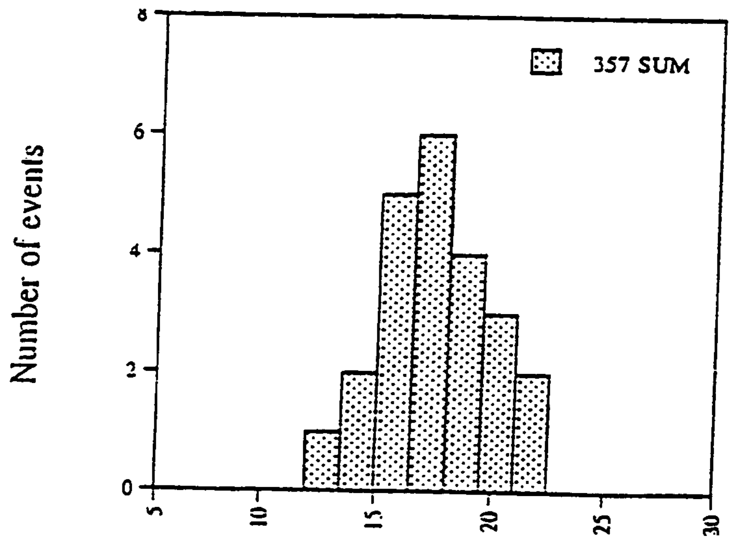
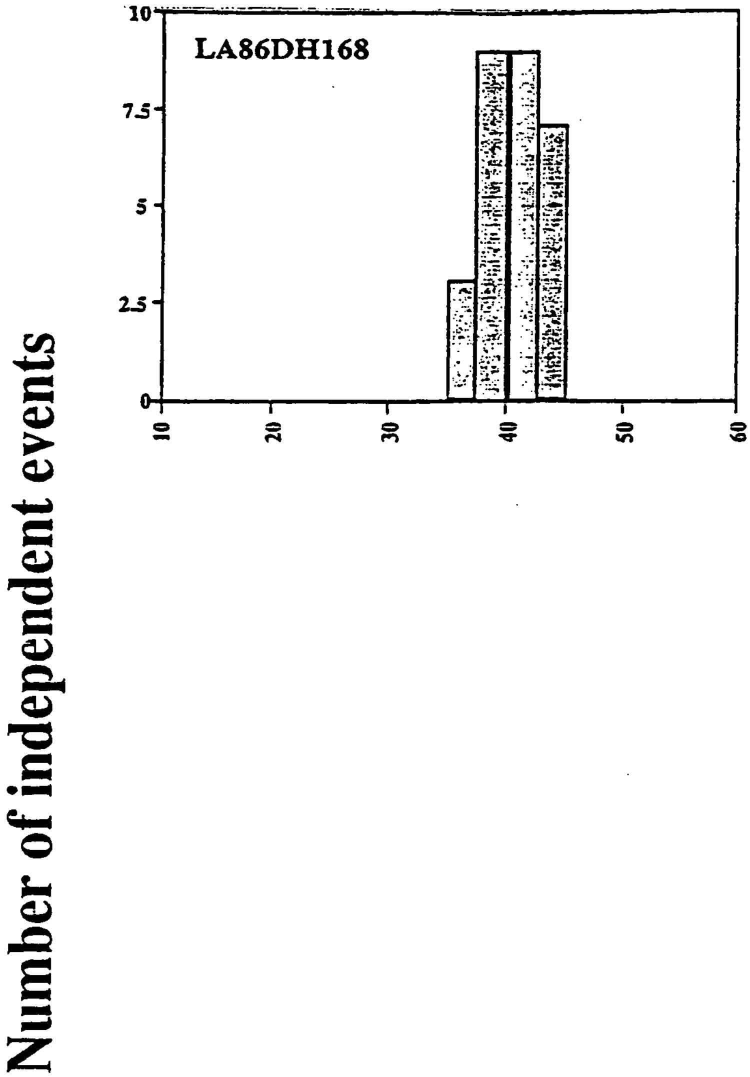


FIGURE 18
2/2

59/166



12:0 levels (w%)

FIGURE 19
1/3

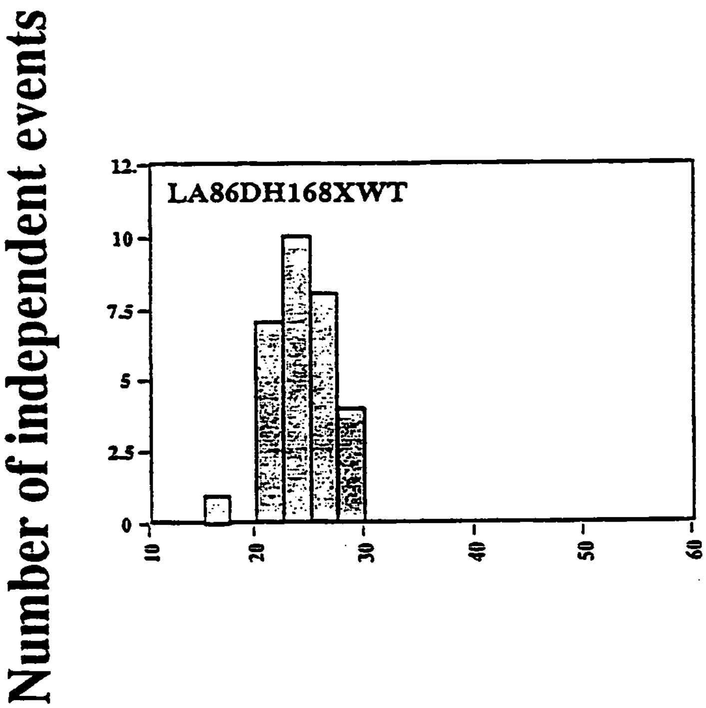

Number of independent events
LA86DH168XWT
12
10
7.5
5
2.5
0
10
20
30
40
50
60

FIGURE 19
3/3

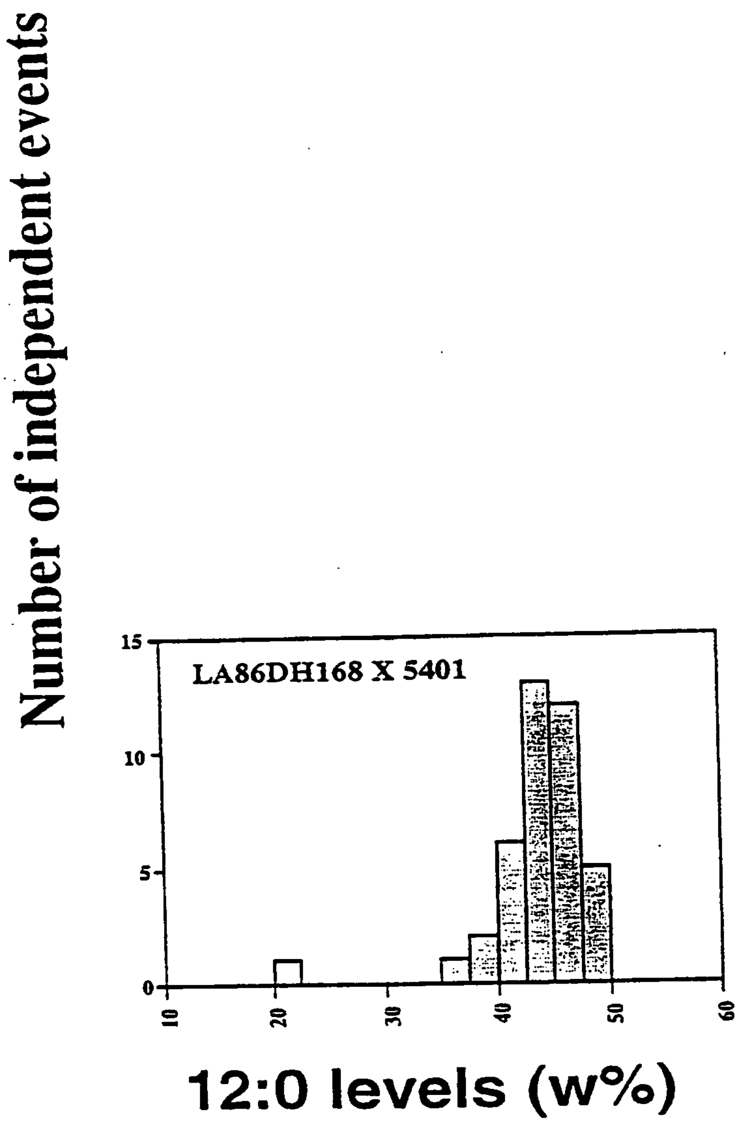

Number of independent events
LA86DH168 X 5401
0
5
10
15
10
20
30
40
50
60
12:0 levels (w%)

FIGURE 19
2/3

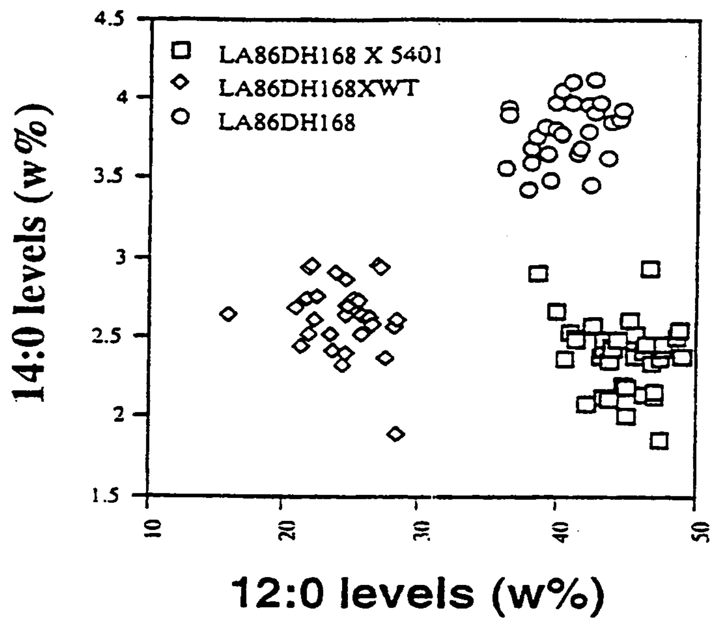
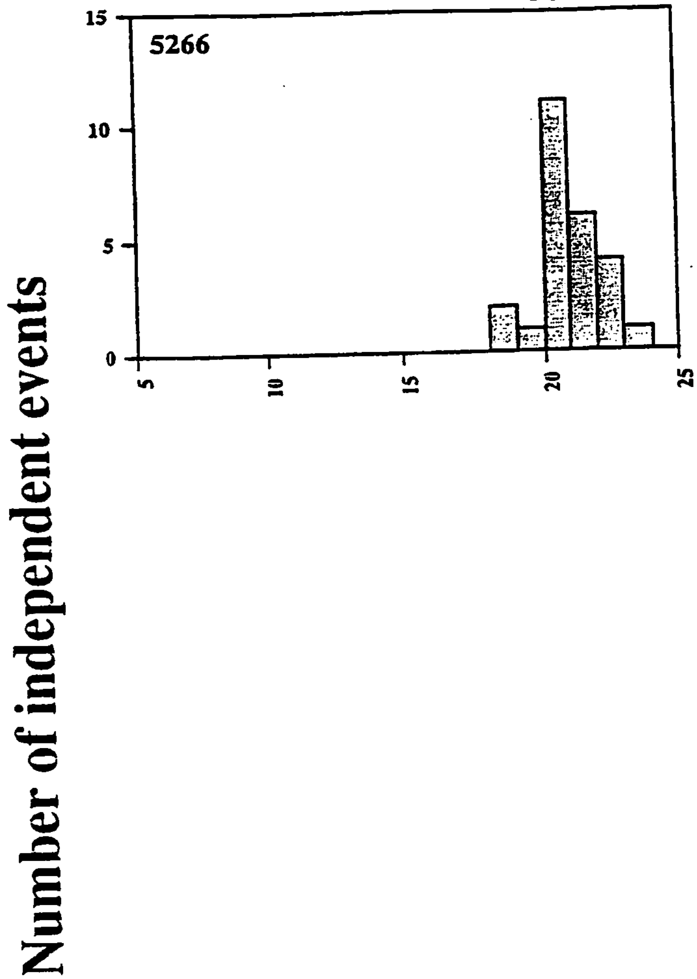

LA86DH168 X 5401
LA86DH168XWT
LA86DH168
14:0 levels (w%)
4.5
4
3.5
3
2.5
2
1.5
10
20
30
40
50
12:0 levels (w%)

FIGURE 20



18:0 levels (w%)

FIGURE -21

1/3

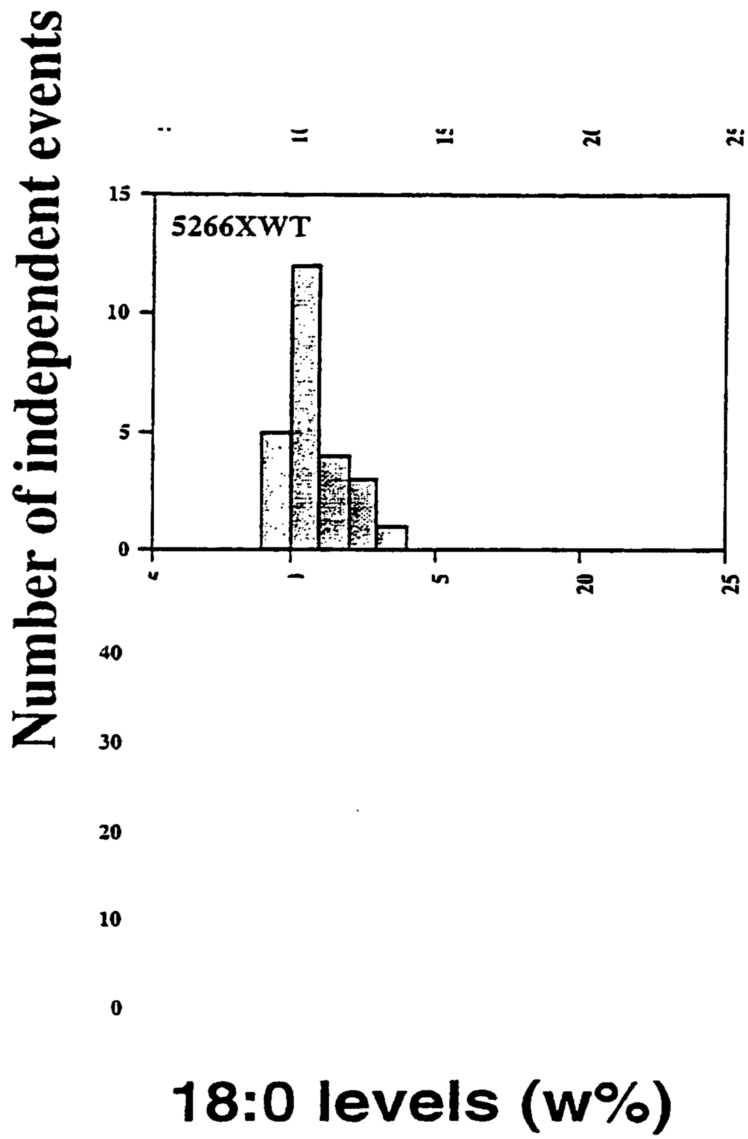

Number of independent events
5266XWT
15
10
5
0
40
30
20
10
0
18:0 levels (w%)

FIGURE 21
2/3

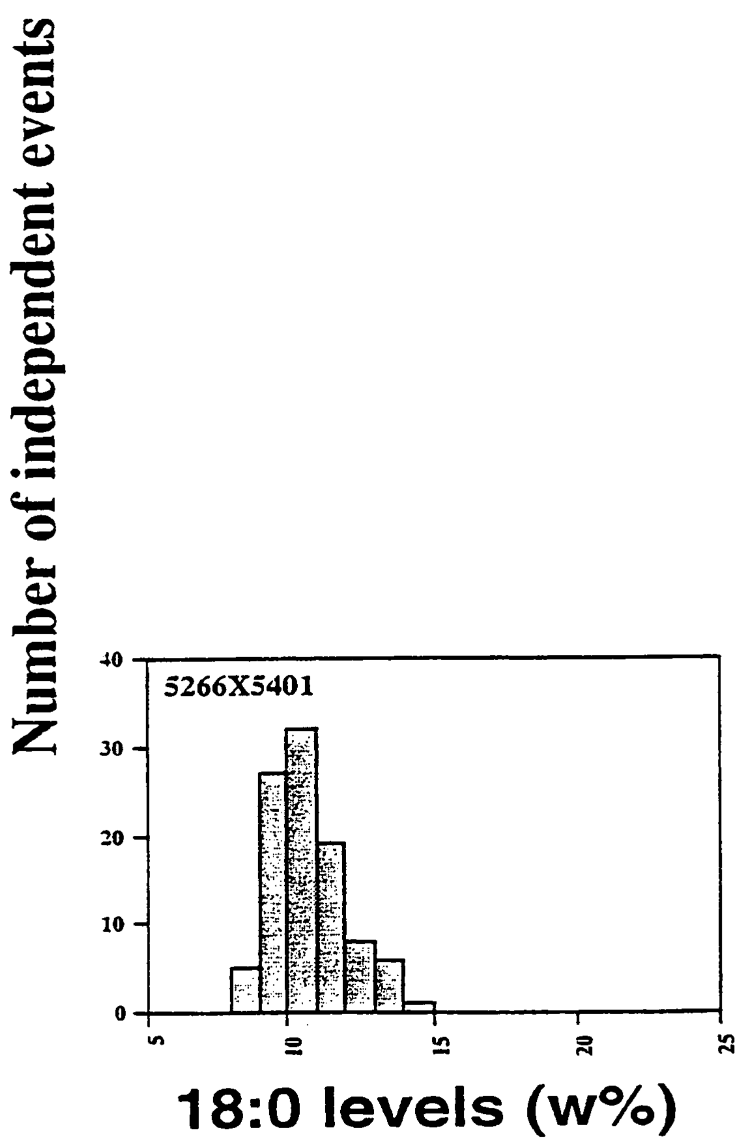

Number of independent events
5266X5401
40
30
20
10
0
5
10
15
20
25
18:0 levels (w%)

FIGURE 21
3/3

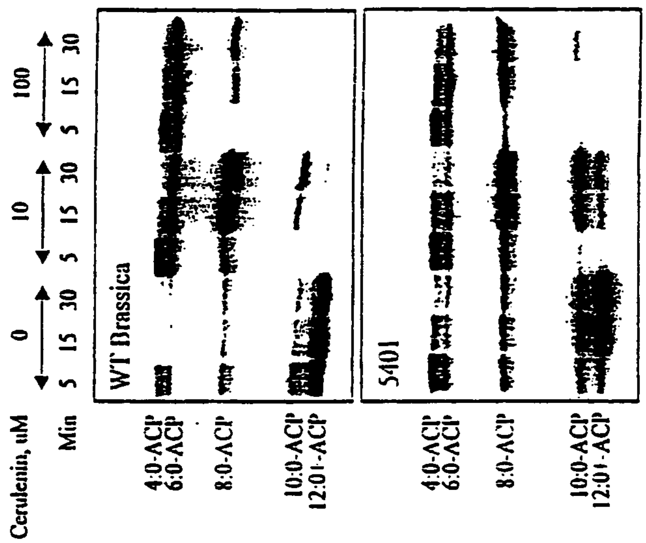

Cerulenin, uM
Min
0
10
100
5 15 30
5 15 30
5 15 30
WT Brassica
4:0-ACP
6:0-ACP
8:0-ACP
10:0-ACP
5401
4:0-ACP
6:0-ACP
8:0-ACP
10:0-ACP

FIGURE 22

CORRECTED VERSION*

28
27

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/53, 15/81, 9/02, 5/10, 1/19, C12P 7/64, C11B 1/00, A61K 31/20, A23L 1/30

A1

(11) International Publication Number: WO 98/46763

(43) International Publication Date: 22 October 1998 (22.10.98)

(21) International Application Number: PCT/US98/07126

(22) International Filing Date: 10 April 1998 (10.04.98)

(30) Priority Data:
08/834,655 11 April 1997 (11.04.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 08/834,655 (CIP)
Filed on 11 April 1997 (11.04.97)

(71) Applicants *(for all designated States except US)*: CALGENE LLC [US/US]; 1920 Fifth Street, Davis, CA 95616 (US). ABBOTT LABORATORIES [US/US]; 100 Abbott Park Road, Abbott Park, IL 60064-3500 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KNUTZON, Deborah [US/US]; 6110 Rockhurst Way, Granite Bay, CA 95746 (US). MUKERJI, Pradip [US/US]; 1069 Arcaro Drive, Gahanna, OH 43230 (US). HUANG, Yung-Sheng [CA/US]; 2462 Danvers Court, Upper Arlington, OH 43220 (US). THURMOND, Jennifer [US/US]; 3702 Adirondack, Columbus, OH 43231 (US). CHAUDHARY, Sunita [IN/US]; 3419 Woodbine Place, Pearland, TX 77584 (US). LEONARD, Amanda, Eun-Yeong [US/US]; 581 Shadewood Court, Gahanna, OH 43230 (US).

(74) Agents: WARD, Michael, R. et al.; Limbach & Limbach L.L.P., 2001 Ferry Building, San Francisco, CA 94111-4262 (US).

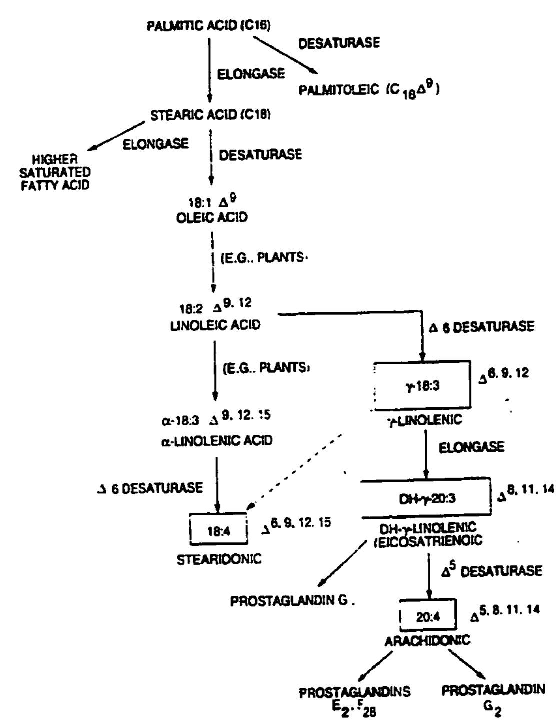
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHODS AND COMPOSITIONS FOR SYNTHESIS OF LONG CHAIN POLYUNSATURATED FATTY ACIDS

(57) Abstract

The present invention relates to fatty acid desaturases able to catalyze the conversion of oleic acid to linoleic acid, linoleic acid to γ-linolenic acid, or of alpha-linolenic acid to stearidonic acid. Nucleic acid sequences encoding desaturases, nucleic acid sequences which hybridize thereto, DNA constructs comprising a desaturase gene, and recombinant host microorganism or animal expressing increased levels of a desaturase are described. Methods for desaturating a fatty acid and for producing a desaturated fatty acid by expressing increased levels of a desaturase are disclosed. Fatty acids, and oils containing them, which have been desaturated by a desaturase produced by recombinant host microorganisms or animals are provided. Pharmaceutical compositions, infant formulas or dietary supplements containing fatty acids which have been desaturated by a desaturase produced by a recombinant host microorganism or animal also are described.



*(Referred to in PCT Gazette No. 11/1999, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

METHODS AND COMPOSITIONS FOR SYNTHESIS OF LONG CHAIN POLYUNSATURATED FATTY ACIDS

RELATED APPLICATIONS

This application is a contiuation-in-part application of United States Patent Application Serial No. 08/834,655 filed April 11, 1997.

INTRODUCTION

Field of the Invention

This invention relates to modulating levels of enzymes and/or enzyme components relating to production of long chain poly-unsaturated fatty acids (PUFAs) in a microorganism or animal.

Background

Two main families of polyunsaturated fatty acids (PUFAs) are the ω3 fatty acids, exemplified by eicosapentaenoic acid (EPA), and the ω6 fatty acids, exemplified by arachidonic acid (ARA). PUFAs are important components of the plasma membrane of the cell, where they may be found in such forms as phospholipids. PUFAs are necessary for proper development, particularly in the developing infant brain, and for tissue formation and repair. PUFAs also serve as precursors to other molecules of importance in human beings and animals, including the prostacyclins, eicosanoids, leukotrienes and prostaglandins. Four major long chain PUFAs of importance include docosahexaenoic acid (DHA) and EPA, which are primarily found in different types of fish oil, γ-linolenic acid (GLA), which is found in the seeds of a number of plants, including evening primrose (*Oenothera biennis*), borage (*Borago officinalis*) and black currants (*Ribes nigrum*), and stearidonic acid (SDA), which is found in marine oils and plant seeds. Both GLA and another important long chain PUFA, arachidonic acid (ARA), are found in filamentous fungi. ARA can be purified from animal tissues including liver and adrenal gland. GLA, ARA, EPA and

SDA are themselves, or are dietary precursors to, important long chain fatty acids involved in prostaglandin synthesis, in treatment of heart disease, and in development of brain tissue.

For DHA, a number of sources exist for commercial production including a variety of marine organisms, oils obtained from cold water marine fish, and egg yolk fractions. For ARA, microorganisms including the genera *Mortierella, Entomophthora, Phytium* and *Porphyridium* can be used for commercial production. Commercial sources of SDA include the genera *Trichodesma and Echium*. Commercial sources of GLA include evening primrose, black currants and borage. However, there are several disadvantages associated with commercial production of PUFAs from natural sources. Natural sources of PUFAs, such as animals and plants, tend to have highly heterogeneous oil compositions. The oils obtained from these sources therefore can require extensive purification to separate out one or more desired PUFAs or to produce an oil which is enriched in one or more PUFA. Natural sources also are subject to uncontrollable fluctuations in availability. Fish stocks may undergo natural variation or may be depleted by overfishing. Fish oils have unpleasant tastes and odors, which may be impossible to economically separate from the desired product, and can render such products unacceptable as food supplements. Animal oils, and particularly fish oils, can accumulate environmental pollutants. Weather and disease can cause fluctuation in yields from both fish and plant sources. Cropland available for production of alternate oil-producing crops is subject to competition from the steady expansion of human populations and the associated increased need for food production on the remaining arable land. Crops which do produce PUFAs, such as borage, have not been adapted to commercial growth and may not perform well in monoculture. Growth of such crops is thus not economically competitive where more profitable and better established crops can be grown. Large scale fermentation of organisms such as *Mortierella* is also expensive. Natural animal tissues contain low amounts of ARA and are difficult to process. Microorganisms such as *Porphyridium* and *Mortierella* are difficult to cultivate on a commercial scale.

Dietary supplements and pharmaceutical formulations containing PUFAs can retain the disadvantages of the PUFA source. Supplements such as fish oil capsules can contain low levels of the particular desired component and thus require large dosages. High dosages result in ingestion of high levels of undesired components, including contaminants. Unpleasant tastes and odors of the supplements can make such regimens undesirable, and may inhibit compliance by the patient. Care must be taken in providing fatty acid supplements, as overaddition may result in suppression of endogenous biosynthetic pathways and lead to competition with other necessary fatty acids in various lipid fractions *in vivo*, leading to undesirable results. For example, Eskimos having a diet high in ω3 fatty acids have an increased tendency to bleed (U.S. Pat. No. 4,874,603).

A number of enzymes are involved in PUFA biosynthesis. Linoleic acid (LA, 18:2 Δ9, 12) is produced from oleic acid (18:1 Δ9) by a Δ12-desaturase. GLA (18:3 Δ6, 9, 12) is produced from linoleic acid (LA, 18:2 Δ9, 12) by a Δ6-desaturase. ARA (20:4 Δ5, 8, 11, 14) production from dihomo-γ-linolenic acid (DGLA, 20:3 Δ8, 11, 14) is catalyzed by a Δ5-desaturase. However, animals cannot desaturate beyond the Δ9 position and therefore cannot convert oleic acid (18:1 Δ9) into linoleic acid (18:2 Δ9, 12). Likewise, α-linolenic acid (ALA, 18:3 Δ9, 12, 15) cannot be synthesized by mammals. Other eukaryotes, including fungi and plants, have enzymes which desaturate at positions Δ12 and Δ15. The major poly-unsaturated fatty acids of animals therefore are either derived from diet and/or from desaturation and elongation of linoleic acid (18:2 Δ9, 12) or α-linolenic acid (18:3 Δ9, 12, 15). Therefore it is of interest to obtain genetic material involved in PUFA biosynthesis from species that naturally produce these fatty acids and to express the isolated material in a microbial or animal system which can be manipulated to provide production of commercial quantities of one or more PUFAs. Thus there is a need for fatty acid desaturases, genes encoding them, and recombinant methods of producing them. A need further exists for oils containing higher relative proportions of and/or

enriched in specific PUFAs. A need also exists for reliable economical methods of producing specific PUFAs.

Relevant Literature

Production of γ-linolenic acid by a Δ6-desaturase is described in USPN 5,552,306. Production of 8, 11-eicosadienoic acid using *Mortierella alpina* is disclosed in USPN 5,376,541. Production of docosahexaenoic acid by dinoflagellates is described in USPN 5,407,957. Cloning of a Δ6-palmitoyl-acyl carrier protein desaturase is described in PCT publication WO 96/13591 and USPN 5,614,400. Cloning of a Δ6-desaturase from borage is described in PCT publication WO 96/21022. Cloning of Δ9-desaturases is described in the published patent applications PCT WO 91/13972, EP 0 550 162 A1, EP 0 561 569 A2, EP 0 644 263 A2, and EP 0 736 598 A1, and in USPN 5,057,419. Cloning of Δ12-desaturases from various organisms is described in PCT publication WO 94/11516 and USPN 5,443,974. Cloning of Δ15-desaturases from various organisms is described in PCT publication WO 93/11245. All publications and U.S. patents or applications referred to herein are hereby incorporated in their entirety by reference.

SUMMARY OF THE INVENTION

Novel compositions and methods are provided for preparation of poly-unsaturated long chain fatty acids. The compositions include nucleic acid encoding a Δ6- and Δ12- desaturase and/or polypeptides having Δ6- and/or Δ12-desaturase activity, the polypeptides, and probes isolating and detecting the same. The methods involve growing a host microorganism or animal expressing an introduced gene or genes encoding at least one desaturase, particularly a Δ6-, Δ9-, Δ12- or Δ15-desaturase. The methods also involve the use of antisense constructs or gene disruptions to decrease or eliminate the expression level of undesired desaturases. Regulation of expression of the desaturase polypeptide(s) provides for a relative increase in desired desaturated PUFAs as a result of altered concentrations of enzymes and substrates involved

in PUFA biosynthesis. The invention finds use, for example, in the large scale production of GLA, DGLA, ARA, EPA, DHA and SDA.

In a preferred embodiment of the invention, an isolated nucleic acid comprising: a nucleotide sequence depicted in Figure 3A-E (SEQ ID NO: 1) or Figure 5A-D (SEQ ID NO: 3), a polypeptide encoded by a nucleotide sequence according Figure 3A-E (SEQ ID NO: 1) or Figure 5A-D (SEQ ID NO: 3), and a purified or isolated polypeptide comprising an amino acid sequence depicted in Figure 3A-E (SEQ ID NO: 2) or Figure 5A-D (SEQ ID NO: 4). In another embodiment of the invention, provided is an isolated nucleic acid encoding a polypeptide having an amino acid sequence depicted in Figure 3A-E (SEQ ID NO: 2) or Figure 5A-D (SEQ ID NO: 4).

Also provided is an isolated nucleic acid comprising a nucleotide sequence which encodes a polypeptide which desaturates a fatty acid molecule at carbon 6 or 12 from the carboxyl end, wherein said nucleotide sequence has an average A/T content of less than about 60%. In a preferred embodiment, the isolated nucleic acid is derived from a fungus, such as a fungus of the genus *Mortierella*. More preferred is a fungus of the species *Mortierella alpina*.

In another preferred embodiment of the invention, an isolated nucleic acid is provided wherein the nucleotide sequence of the nucleic acid is depicted in Figure 3A-E (SEQ ID NO: 1) or Figure 5A-D (SEQ ID NO: 3). The invention also provides an isolated or purified polypeptide which desaturates a fatty acid molecule at carbon 6 or 12 from the carboxyl end, wherein the polypeptide is a eukaryotic polypeptide or is derived from a eukaryotic polypeptide, where a preferred eukaryotic polypeptide is derived from a fungus.

The present invention further includes a nucleic acid sequence which hybridizes to Figure 3A-E (SEQ ID NO: 1) or Figure 5A-D (SEQ ID NO: 3). Preferred is an isolated nucleic acid having a nucleotide sequence with at least about 50% homology to Figure 3A-E (SEQ ID NO: 1) or Figure 5A-D (SEQ ID NO: 3). The invention also includes an isolated nucleic acid having a nucleotide sequence with at least about 50% homology to Figure 3A-E (SEQ ID NO: 1) or Figure 5A-D (SEQ ID NO: 3). In a preferred embodiment, the

nucleic acid of the invention includes a nucleotide sequence which encodes an amino acid sequence depicted in Figure 3A-D (SEQ ID NO: 2) which is selected from the group consisting of amino acid residues 50-53, 39-43, 172-176, 204-213, and 390-402.

Also provided by the present invention is a nucleic acid construct comprising a nucleotide sequence depicted in a Figure 3A-E (SEQ ID NO: 1) or Figure 5A-D (SEQ ID NO: 3) linked to a heterologous nucleic acid. In another embodiment, a nucleic acid construct is provided which comprises a nucleotide sequence depicted in a Figure 3A-E (SEQ ID NO: 1) or Figure 5A-D (SEQ ID NO: 3) operably associated with an expression control sequence functional in a host cell. The host cell is either eukaryotic or prokaryotic. Preferred eukaryotic host cells are those selected from the group consisting of a mammalian cell, an insect cell, a fungal cell, and an algae cell. Preferred mammalian cells include an avian cell, a preferred fungal cell includes a yeast cell, and a preferred algae cell is a marine algae cell. Preferred prokaryotic cells include those selected from the group consisting of a bacteria, a cyanobacteria, cells which contain a bacteriophage, and/or a virus. The DNA sequence of the recombinant host cell preferably contains a promoter which is functional in the host cell, which promoter is preferably inducible. In a more preferred embodiment, the microbial cell is a fungal cell of the genus *Mortierella*, with a more preferred fungus is of the species *Mortierella alpina*.

In addition, the present invention provides a nucleic acid construct comprising a nucleotide sequence which encodes a polypeptide comprising an amino acid sequence which corresponds to or is complementary to an amino acid sequence depicted in Figure 3A-E (SEQ ID NO: 2) or Figure 5A-D (SEQ ID NO: 4), wherein the nucleic acid is operably associated with an expression control sequence functional in a microbial cell, wherein the nucleotide sequence encodes a functionally active polypeptide which desaturates a fatty acid molecule at carbon 6 or carbon 12 from the carboxyl end of a fatty acid molecule. Another embodiment of the present invention is a nucleic acid construct comprising a nucleotide sequence which encodes a functionally active Δ6-desaturase having an amino acid sequence which corresponds to or is

complementary to all of or a portion of an amino acid sequence depicted in a Figure 3A-E (SEQ ID NO: 2), wherein the nucleotide sequence is operably associated with a transcription control sequence functional in a host cell.

Yet another embodiment of the present invention is a nucleic acid construct comprising a nucleotide sequence which encodes a functionally active Δ12-desaturase having an amino acid sequence which corresponds to or is complementary to all of or a portion of an amino acid sequence depicted in a Figure 5A-D (SEQ ID NO: 4), wherein the nucleotide sequence is operably associated with a transcription control sequence functional in a host cell. The host cell, is either a eukaryotic or prokaryotic host cell. Preferred eukaryotic host cells are those selected from the group consisting of a mammalian cell, an insect cell, a fungal cell, and an algae cell. Preferred mammalian cells include an avian cell, a preferred fungal cell includes a yeast cell, and a preferred algae cell is a marine algae cell. Preferred prokaryotic cells include those selected from the group consisting of a bacteria, a cyanobacteria, cells which contain a bacteriophage, and/or a virus. The DNA sequence of the recombinant host cell preferably contains a promoter which is functional in the host cell and which preferably is inducible. A preferred recombinant host cell is a microbial cell such as a yeast cell, such as a *Saccharomyces* cell.

The present invention also provides a recombinant microbial cell comprising at least one copy of a nucleic acid which encodes a functionally active *Mortierella alpina* fatty acid desaturase having an amino acid sequence as depicted in Figure 3A-E (SEQ ID NO: 2), wherein the cell or a parent of the cell was transformed with a vector comprising said DNA sequence, and wherein the DNA sequence is operably associated with an expression control sequence. In a preferred embodiment, the cell is a microbial cell which is enriched in 18:2 fatty acids, particularly where the microbial cell is from a genus selected from the group consisting of a prokaryotic cell and eukaryotic cell. In another preferred embodiment, the microbial cell according to the invention includes an expression control sequence which is endogenous to the microbial cell.

Also provided by the present invention is a method for production of GLA in a host cell, where the method comprises growing a host culture having a plurality of host cells which contain one or more nucleic acids encoding a polypeptide which converts LA to GLA, wherein said one or more nucleic acids is operably associated with an expression control sequence, under conditions whereby said one or more nucleic acids are expressed, whereby GLA is produced in the host cell. In several preferred embodiments of the methods, the polypeptide employed in the method is a functionally active enzyme which desaturates a fatty acid molecule at carbon 6 from the carboxyl end of a fatty acid molecule; the said one or more nucleic acids is derived from a *Mortierella alpina*; the substrate for the polypeptide is exogenously supplied; the host cells are microbial cells; the microbial cells are yeast cells, such as *Saccharomyces* cells; and the growing conditions are inducible.

Also provided is an oil comprising one or more PUFA, wherein the amount of said one or more PUFAs is approximately 0.3-30% arachidonic acid (ARA), approximately 0.2-30% dihomo-γ-linolenic acid (DGLA), and approximately 0.2-30% γ-linoleic acid (GLA). A preferred oil of the invention is one in which the ratio of ARA:DGLA:GLA is approximately 1.0:19.0:30 to 6.0:1.0:0.2. Another preferred embodiment of the invention is a pharmaceutical composition comprising the oils in a pharmaceutically acceptable carrier. Further provided is a nutritional composition comprising the oils of the invention. The nutritional compositions of the invention preferably are administered to a mammalian host parenterally or internally. A preferred composition of the invention for internal consumption is an infant formula. In a preferred embodiment, the nutritional compositions of the invention are in a liquid form or a solid form, and can be formulated in or as a dietary supplement, and the oils provided in encapsulated form. The oils of the invention can be free of particular components of other oils and can be derived from a microbial cell, such as a yeast cell.

The present invention further provides a method for desaturating a fatty acid. In a preferred embodiment the method comprises culturing a recombinant microbial cell according to the invention under conditions suitable for

expression of a polypeptide encoded by said nucleic acid, wherein the host cell further comprises a fatty acid substrate of said polypeptide. Also provided is a fatty acid desaturated by such a method, and an oil composition comprising a fatty acid produced according to the methods of the invention.

The present invention further includes a purified nucleotide sequence or polypeptide sequence that is substantially related or homologous to the nucleotide and peptide sequences presented in SEQ ID NO:1 - SEQ ID NO:40. The present invention is further directed to methods of using the sequences presented in SEQ ID NO:1 to SEQ ID NO:40 as probes to identify related sequences, as components of expression systems and as components of systems useful for producing transgenic oil.

The present invention is further directed to formulas, dietary supplements or dietary supplements in the form of a liquid or a solid containing the long chain fatty acids of the invention. These formulas and supplements may be administered to a human or an animal.

The formulas and supplements of the invention may further comprise at least one macronutrient selected from the group consisting of coconut oil, soy oil, canola oil, mono- and diglycerides, glucose, edible lactose, electrodialysed whey, electrodialysed skim milk, milk whey, soy protein, and other protein hydrolysates.

The formulas of the present invention may further include at least one vitamin selected from the group consisting of Vitamins A, C, D, E, and B complex; and at least one mineral selected from the group consisting of calcium, magnesium, zinc, manganese, sodium, potassium, phosphorus, copper, chloride, iodine, selenium, and iron.

The present invention is further directed to a method of treating a patient having a condition caused by insuffient intake or production of polyunsaturated fatty acids comprising administering to the patient a dietary substitute of the invention in an amount sufficient to effect treatment of the patient.

The present invention is further directed to cosmetic and pharmaceutical compositions of the material of the invention.

The present invention is further directed to transgenic oils in pharmaceutically acceptable carriers. The present invention is further directed to nutritional supplements, cosmetic agents and infant formulae containing transgenic oils.

The present invention is further directed to a method for obtaining altered long chain polyunsaturated fatty acid biosynthesis comprising the steps of: growing a microbe having cells which contain a transgene which encodes a transgene expression product which desaturates a fatty acid molecule at carbon 6 or 12 from the carboxyl end of said fatty acid molecule, wherein the trangene is operably associated with an expression control sequence, under conditions whereby the transgene is expressed, whereby long chain polyunsaturated fatty acid biosynthesis in the cells is altered.

The present invention is further directed toward pharmaceutical compositions comprising at least one nutrient selected from the group consisting of a vitamin, a mineral, a carbohydrate, a sugar, an amino acid, a free fatty acid, a phospholipid, an antioxidant, and a phenolic compound.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows possible pathways for the synthesis of arachidonic acid (20:4 Δ5, 8, 11, 14) and stearidonic acid (18:4 Δ6, 9, 12, 15) from palmitic acid (C_{16}) from a variety of organisms, including algae, *Mortierella* and humans. These PUFAs can serve as precursors to other molecules important for humans and other animals, including prostacyclins, leukotrienes, and prostaglandins, some of which are shown.

Figure 2 shows possible pathways for production of PUFAs in addition to ARA, including EPA and DHA, again compiled from a variety of organisms.

Figure 3A-E shows the DNA sequence of the *Mortierella alpina* Δ6-desaturase and the deduced amino acid sequence:

Figure 3A-E (SEQ ID NO 1 Δ6 DESATURASE cDNA)

Figure 3A-E (SEQ ID NO 2 Δ6 DESATURASE AMINO ACID)

Figure 4 shows an alignment of a portion of the *Mortierella alpina* Δ6-desaturase amino acid sequence with other related sequences.

Figure 5A-D shows the DNA sequence of the *Mortierella alpina* Δ12-desaturase and the deduced amino acid sequence:

Figure 5A-D (SEQ ID NO 3 Δ12 DESATURASE cDNA)

Figure 5A-D (SEQ ID NO 4 Δ12 DESATURASE AMINO ACID).

Figures 6A and 6B show the effect of different expression constructs on expression of GLA in yeast.

Figures 7A and 7B show the effect of host strain on GLA production.

Figures 8A and 8B show the effect of temperature on GLA production in *S. cerevisiae* strain SC334.

Figure 9 shows alignments of the protein sequence of the Ma 29 and contig 253538a.

Figure 10 shows alignments of the protein sequence of Ma 524 and contig 253538a.

BRIEF DESCRIPTION OF THE SEQUENCE LISTINGS

SEQ ID NO:1 shows the DNA sequence of the *Mortierella alpina* Δ6-desaturase.

SEQ ID NO:2 shows the protein sequence of the *Mortierella alpina* Δ6-desaturase.

SEQ ID NO:3 shows the DNA sequence of the *Mortierella alpina* Δ12-desaturase.

SEQ ID NO:4 shows the protein sequence of the *Mortierella alpina* Δ12-desaturase.

SEQ ID NO:5-11 show various desaturase sequences.

SEQ ID NO:13-18 show various PCR primer sequences.

SEQ ID NO:19 and SEQ ID NO:20 show the nucleotide and amino acid sequence of a *Dictyostelium discoideum* desaturase.

SEQ ID NO:21 and SEQ ID NO:22 show the nucleotide and amino acid sequence of a *Phaeodactylum tricornutum* desaturase.

SEQ ID NO:23-26 show the nucleotide and deduced amino acid sequence of a *Schizochytrium* cDNA clone.

SEQ ID NO: 27-33 show nucleotide sequences for human desaturases.

SEQ ID NO:34 - SEQ ID NO:40 show peptide sequences for human desaturases.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In order to ensure a complete understanding of the invention, the following definitions are provided:

Δ5-Desaturase: Δ5 desaturase is an enzyme which introduces a double bond between carbons 5 and 6 from the carboxyl end of a fatty acid molecule.

Δ6-Desaturase: Δ6-desaturase is an enzyme which introduces a double bond between carbons 6 and 7 from the carboxyl end of a fatty acid molecule.

Δ9-Desaturase: Δ9-desaturase is an enzyme which introduces a double bond between carbons 9 and 10 from the carboxyl end of a fatty acid molecule.

Δ12-Desaturase: Δ12-desaturase is an enzyme which introduces a double bond between carbons 12 and 13 from the carboxyl end of a fatty acid molecule.

Fatty Acids: Fatty acids are a class of compounds containing a long hydrocarbon chain and a terminal carboxylate group. Fatty acids include the following:

Fatty Acid		
12:0	lauric acid	
16:0	palmitic acid	

Fatty Acid		
16:1	palmitoleic acid	
18:0	stearic acid	
18:1	oleic acid	Δ9-18:1
18:2 Δ5,9	taxoleic acid	Δ5,9-18:2
18:2 Δ6,9	6,9-octadecadienoic acid	Δ6,9-18:2
18:2	Linolenic acid	Δ9,12-18:2 (LA)
18:3 Δ6,9,12	Gamma-linolenic acid	Δ6,9,12-18:3 (GLA)
18:3 Δ5,9,12	Pinolenic acid	Δ5,9,12-18:3
18:3	alpha-linoleic acid	Δ9,12,15-18:3 (ALA)
18:4	stearidonic acid	Δ6,9,12,15-18:4 (SDA)
20:0	Arachidic acid	
20:1	Eicoscenic Acid	
22:0	behehic acid	
22:1	erucic acid	
22:2	docasadienoic acid	
20:4 ω6	arachidonic acid	Δ5,8,11,14-20:4 (ARA)
20:3 ω6	ω6-eicosatrienoic dihomo-gamma linolenic	Δ8,11,14-20:3 (DGLA)
20:5 ω3	Eicosapentanoic (Timnodonic acid)	Δ5,8,11,14,17-20:5 (EPA)
20:3 ω3	ω3-eicosatrienoic	Δ11,16,17-20:3
20:4 ω3	ω3-eicosatetraenoic	Δ8,11,14,17-20:4
22:5 ω3	Docosapentaenoic	Δ7,10,13,16,19-22:5 (ω3DPA)
22:6 ω3	Docosahexaenoic (cervonic acid)	Δ4,7,10,13,16,19-22:6 (DHA)
24:0	Lignoceric acid	

Taking into account these definitions, the present invention is directed to novel DNA sequences, DNA constructs, methods and compositions are provided which permit modification of the poly-unsaturated long chain fatty acid content of, for example, microbial cells or animals. Host cells are manipulated to express a sense or antisense transcript of a DNA encoding a polypeptide(s) which catalyzes the desaturation of a fatty acid. The substrate(s) for the expressed enzyme may be produced by the host cell or may be exogenously supplied. To achieve expression, the transformed DNA is

operably associated with transcriptional and translational initiation and termination regulatory regions that are functional in the host cell. Constructs comprising the gene to be expressed can provide for integration into the genome of the host cell or can autonomously replicate in the host cell. For production of linoleic acid (LA), the expression cassettes generally used include a cassette which provides for Δ12-desaturase activity, particularly in a host cell which produces or can take up oleic acid (U.S. Patent No. 5,443,974). Production of LA also can be increased by providing an expression cassette for a Δ9-desaturase where that enzymatic activity is limiting. For production of ALA, the expression cassettes generally used include a cassette which provides for Δ15- or ω3-desaturase activity, particularly in a host cell which produces or can take up LA. For production of GLA or SDA, the expression cassettes generally used include a cassette which provides for Δ6-desaturase activity, particularly in a host cell which produces or can take up LA or ALA, respectively. Production of ω6-type unsaturated fatty acids, such as LA or GLA, is favored in a host microorganism or animal which is incapable of producing ALA. The host ALA production can be removed, reduced and/or inhibited by inhibiting the activity of a Δ15- or ω3- type desaturase (see Figure 2). This can be accomplished by standard selection, providing an expression cassette for an antisense Δ15 or ω3 transcript, by disrupting a target Δ15- or ω3-desaturase gene through insertion, deletion, substitution of part or all of the target gene, or by adding an inhibitor of Δ15- or ω3-desaturase. Similarly, production of LA or ALA is favored in a microorganism or animal having Δ6-desaturase activity by providing an expression cassette for an antisense Δ6 transcript, by disrupting a Δ6-desaturase gene, or by use of a Δ6-desaturase inhibitor.

MICROBIAL PRODUCTION OF FATTY ACIDS

Microbial production of fatty acids has several advantages over purification from natural sources such as fish or plants. Many microbes are known with greatly simplified oil compositions compared with those of higher organisms, making purification of desired components easier. Microbial production is not subject to fluctuations caused by external variables such as

weather and food supply. Microbially produced oil is substantially free of contamination by environmental pollutants. Additionally, microbes can provide PUFAs in particular forms which may have specific uses. For example, *Spirulina* can provide PUFAs predominantly at the first and third positions of triglycerides; digestion by pancreatic lipases preferentially releases fatty acids from these positions. Following human or animal ingestion of triglycerides derived from *Spirulina*, these PUFAs are released by pancreatic lipases as free fatty acids and thus are directly available, for example, for infant brain development. Additionally, microbial oil production can be manipulated by controlling culture conditions, notably by providing particular substrates for microbially expressed enzymes, or by addition of compounds which suppress undesired biochemical pathways. In addition to these advantages, production of fatty acids from recombinant microbes provides the ability to alter the naturally occurring microbial fatty acid profile by providing new synthetic pathways in the host or by suppressing undesired pathways, thereby increasing levels of desired PUFAs, or conjugated forms thereof, and decreasing levels of undesired PUFAs.

PRODUCTION OF FATTY ACIDS IN ANIMALS

Production of fatty acids in animals also presents several advantages. Expression of desaturase genes in animals can produce greatly increased levels of desired PUFAs in animal tissues, making recovery from those tissues more economical. For example, where the desired PUFAs are expressed in the breast milk of animals, methods of isolating PUFAs from animal milk are well established. In addition to providing a source for purification of desired PUFAs, animal breast milk can be manipulated through expression of desaturase genes, either alone or in combination with other human genes, to provide animal milks substantially similar to human breast milk during the different stages of infant development. Humanized animal milks could serve as infant formulas where human nursing is impossible or undesired, or in cases of malnourishment or disease.

Depending upon the host cell, the availability of substrate, and the desired end product(s), several polypeptides, particularly desaturases, are of

interest. By "desaturase" is intended a polypeptide which can desaturate one or more fatty acids to produce a mono- or poly-unsaturated fatty acid or precursor thereof of interest. Of particular interest are polypeptides which can catalyze the conversion of stearic acid to oleic acid, of oleic acid to LA, of LA to ALA, of LA to GLA, or of ALA to SDA, which includes enzymes which desaturate at the Δ9, Δ12, (ω6), Δ15, (ω3) or Δ6 positions. By "polypeptide" is meant any chain of amino acids, regardless of length or post-translational modification, for example, glycosylation or phosphorylation. Considerations for choosing a specific polypeptide having desaturase activity include the pH optimum of the polypeptide, whether the polypeptide is a rate limiting enzyme or a component thereof, whether the desaturase used is essential for synthesis of a desired poly-unsaturated fatty acid, and/or co-factors required by the polypeptide. The expressed polypeptide preferably has parameters compatible with the biochemical environment of its location in the host cell. For example, the polypeptide may have to compete for substrate with other enzymes in the host cell. Analyses of the K_m and specific activity of the polypeptide in question therefore are considered in determining the suitability of a given polypeptide for modifying PUFA production in a given host cell. The polypeptide used in a particular situation is one which can function under the conditions present in the intended host cell but otherwise can be any polypeptide having desaturase activity which has the desired characteristic of being capable of modifying the relative production of a desired PUFA.

For production of linoleic acid from oleic acid, the DNA sequence used encodes a polypeptide having Δ12-desaturase activity. For production of GLA from linoleic acid, the DNA sequence used encodes a polypeptide having Δ6-desaturase activity. In particular instances, expression of Δ6-desaturase activity can be coupled with expression of Δ12-desaturase activity and the host cell can optionally be depleted of any Δ15-desaturase activity present, for example by providing a transcription cassette for production of antisense sequences to the Δ15-desaturase transcription product, by disrupting the Δ15-desaturase gene, or by using a host cell which naturally has, or has been mutated to have, low Δ15-desaturase activity. Inhibition of undesired desaturase pathways also can be

accomplished through the use of specific desaturase inhibitors such as those described in U.S. Patent No. 4,778,630. Also, a host cell for Δ6-desaturase expression may have, or have been mutated to have, high Δ12-desaturase activity. The choice of combination of cassettes used depends in part on the PUFA profile and/or desaturase profile of the host cell. Where the host cell expresses Δ12-desaturase activity and lacks or is depleted in Δ15-desaturase activity, overexpression of Δ6-desaturase alone generally is sufficient to provide for enhanced GLA production. Where the host cell expresses Δ9-desaturase activity, expression of a Δ12- and a Δ6-desaturase can provide for enhanced GLA production. When Δ9-desaturase activity is absent or limiting, an expression cassette for Δ9-desaturase can be used. A scheme for the synthesis of arachidonic acid (20:4 $\Delta^{5,8,11,14}$) from stearic acid (18:0) is shown in Figure 2. A key enzyme in this pathway is a Δ6-desaturase which converts the linoleic acid into γ-linolenic acid. Conversion of α-linolenic acid (ALA) to stearidonic acid by a Δ6-desaturase also is shown.

SOURCES OF POLYPEPTIDES HAVING DESATURASE ACTIVITY

A source of polypeptides having desaturase activity and oligonucleotides encoding such polypeptides are organisms which produce a desired polyunsaturated fatty acid. As an example, microorganisms having an ability to produce GLA or ARA can be used as a source of Δ6- or Δ12- desaturase activity. Such microorganisms include, for example, those belonging to the genera *Mortierella, Conidiobolus, Pythium, Phytophathora, Penicillium, Porphyridium, Coidosporium, Mucor, Fusarium, Aspergillus, Rhodotorula,* and *Entomophthora*. Within the genus *Porphyridium*, of particular interest is *Porphyridium cruentum.* Within the genus *Mortierella*, of particular interest are *Mortierella elongata, Mortierella exigua, Mortierella hygrophila, Mortierella ramanniana*, var. angulispora, and *Mortierella alpina*. Within the genus *Mucor*, of particular interest are *Mucor circinelloides* and *Mucor javanicus*.

DNAs encoding desired desaturases can be identified in a variety of ways. As an example, a source of the desired desaturase, for example genomic

or cDNA libraries from *Mortierella*, is screened with detectable enzymatically- or chemically-synthesized probes, which can be made from DNA, RNA, or non-naturally occurring nucleotides, or mixtures thereof. Probes may be enzymatically synthesized from DNAs of known desaturases for normal or reduced-stringency hybridization methods. Oligonucleotide probes also can be used to screen sources and can be based on sequences of known desaturases, including sequences conserved among known desaturases, or on peptide sequences obtained from the desired purified protein. Oligonucleotide probes based on amino acid sequences can be degenerate to encompass the degeneracy of the genetic code, or can be biased in favor of the preferred codons of the source organism. Oligonucleotides also can be used as primers for PCR from reverse transcribed mRNA from a known or suspected source; the PCR product can be the full length cDNA or can be used to generate a probe to obtain the desired full length cDNA. Alternatively, a desired protein can be entirely sequenced and total synthesis of a DNA encoding that polypeptide performed.

Once the desired genomic or cDNA has been isolated, it can be sequenced by known methods. It is recognized in the art that such methods are subject to errors, such that multiple sequencing of the same region is routine and is still expected to lead to measurable rates of mistakes in the resulting deduced sequence, particularly in regions having repeated domains, extensive secondary structure, or unusual base compositions, such as regions with high GC base content. When discrepancies arise, resequencing can be done and can employ special methods. Special methods can include altering sequencing conditions by using: different temperatures; different enzymes; proteins which alter the ability of oligonucleotides to form higher order structures; altered nucleotides such as ITP or methylated dGTP; different gel compositions, for example adding formamide; different primers or primers located at different distances from the problem region; or different templates such as single stranded DNAs. Sequencing of mRNA also can be employed.

For the most part, some or all of the coding sequence for the polypeptide having desaturase activity is from a natural source. In some situations, however, it is desirable to modify all or a portion of the codons, for example, to

enhance expression, by employing host preferred codons. Host preferred codons can be determined from the codons of highest frequency in the proteins expressed in the largest amount in a particular host species of interest. Thus, the coding sequence for a polypeptide having desaturase activity can be synthesized in whole or in part. All or portions of the DNA also can be synthesized to remove any destabilizing sequences or regions of secondary structure which would be present in the transcribed mRNA. All or portions of the DNA also can be synthesized to alter the base composition to one more preferable in the desired host cell. Methods for synthesizing sequences and bringing sequences together are well established in the literature. *In vitro* mutagenesis and selection, site-directed mutagenesis, or other means can be employed to obtain mutations of naturally occurring desaturase genes to produce a polypeptide having desaturase activity *in vivo* with more desirable physical and kinetic parameters for function in the host cell, such as a longer half-life or a higher rate of production of a desired polyunsaturated fatty acid.

Mortieralla alpina **Desaturase**

Of particular interest is the *Mortierella alpina* Δ6-desaturase, which has 457 amino acids and a predicted molecular weight of 51.8 kD; the amino acid sequence is shown in Figure 3. The gene encoding the *Mortierella alpina* Δ6-desaturase can be expressed in transgenic microorganisms or animals to effect greater synthesis of GLA from linoleic acid or of stearidonic acid from ALA. Other DNAs which are substantially identical to the *Mortierella alpina* Δ6-desaturase DNA, or which encode polypeptides which are substantially identical to the *Mortierella alpina* Δ6-desaturase polypeptide, also can be used. By substantially identical is intended an amino acid sequence or nucleic acid sequence exhibiting in order of increasing preference at least 60%, 80%, 90% or 95% homology to the *Mortierella alpina* Δ6-desaturase amino acid sequence or nucleic acid sequence encoding the amino acid sequence. For polypeptides, the length of comparison sequences generally is at least 16 amino acids, preferably at least 20 amino acids, or most preferably 35 amino acids. For nucleic acids, the length of comparison sequences generally is at least 50 nucleotides,

preferably at least 60 nucleotides, and more preferably at least 75 nucleotides, and most preferably, 110 nucleotides. Homology typically is measured using sequence analysis software, for example, the Sequence Analysis software package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wisconsin 53705, MEGAlign (DNAStar, Inc., 1228 S. Park St., Madison, Wisconsin 53715), and MacVector (Oxford Molecular Group, 2105 S. Bascom Avenue, Suite 200, Campbell, California 95008). Such software matches similar sequences by assigning degrees of homology to various substitutions, deletions, and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine and leucine; aspartic acid, glutamic acid, asparagine, and glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine. Substitutions may also be made on the basis of conserved hydrophobicity or hydrophilicity (Kyte and Doolittle, *J. Mol. Biol.* 157: 105-132, 1982), or on the basis of the ability to assume similar polypeptide secondary structure (Chou and Fasman, *Adv. Enzymol.* 47: 45-148, 1978).

Also of interest is the *Mortierella alpina* Δ12-desaturase, the nucleotide and amino acid sequence of which is shown in Figure 5. The gene encoding the *Mortierella alpina* Δ12-desaturase can be expressed in transgenic microorganisms or animals to effect greater synthesis of LA from oleic acid. Other DNAs which are substantially identical to the *Mortierella alpina* Δ12-desaturase DNA, or which encode polypeptides which are substantially identical to the *Mortierella alpina* Δ12-desaturase polypeptide, also can be used.

Other Desaturases

Encompassed by the present invention are related desaturases from the same or other organisms. Such related desaturases include variants of the disclosed Δ6- or Δ12-desaturase naturally occurring within the same or different species of *Mortierella*, as well as homologues of the disclosed Δ6- or Δ12-desaturase from other species. Also included are desaturases which, although

not substantially identical to the *Mortierella alpina* Δ6- or Δ12-desaturase, desaturate a fatty acid molecule at carbon 6 or 12, respectively, from the carboxyl end of a fatty acid molecule, or at carbon 12 or 6 from the terminal methyl carbon in an 18 carbon fatty acid molecule. Related desaturases can be identified by their ability to function substantially the same as the disclosed desaturases; that is, are still able to effectively convert LA to GLA, ALA to SDA or oleic acid to LA. Related desaturases also can be identified by screening sequence databases for sequences homologous to the disclosed desaturases, by hybridization of a probe based on the disclosed desaturases to a library constructed from the source organism, or by RT-PCR using mRNA from the source organism and primers based on the disclosed desaturases. Such desaturases include those from humans, *Dictyostelium discoideum* and *Phaeodactylum tricornum.*

The regions of a desaturase polypeptide important for desaturase activity can be determined through routine mutagenesis, expression of the resulting mutant polypeptides and determination of their activities. Mutants may include deletions, insertions and point mutations, or combinations thereof. A typical functional analysis begins with deletion mutagenesis to determine the N- and C-terminal limits of the protein necessary for function, and then internal deletions, insertions or point mutants are made to further determine regions necessary for function. Other techniques such as cassette mutagenesis or total synthesis also can be used. Deletion mutagenesis is accomplished, for example, by using exonucleases to sequentially remove the 5' or 3' coding regions. Kits are available for such techniques. After deletion, the coding region is completed by ligating oligonucleotides containing start or stop codons to the deleted coding region after 5' or 3' deletion, respectively. Alternatively, oligonucleotides encoding start or stop codons are inserted into the coding region by a variety of methods including site-directed mutagenesis, mutagenic PCR or by ligation onto DNA digested at existing restriction sites. Internal deletions can similarly be made through a variety of methods including the use of existing restriction sites in the DNA, by use of mutagenic primers via site directed mutagenesis or mutagenic PCR. Insertions are made through methods such as linker-scanning

mutagenesis, site-directed mutagenesis or mutagenic PCR. Point mutations are made through techniques such as site-directed mutagenesis or mutagenic PCR.

Chemical mutagenesis also can be used for identifying regions of a desaturase polypeptide important for activity. A mutated construct is expressed, and the ability of the resulting altered protein to function as a desaturase is assayed. Such structure-function analysis can determine which regions may be deleted, which regions tolerate insertions, and which point mutations allow the mutant protein to function in substantially the same way as the native desaturase. All such mutant proteins and nucleotide sequences encoding them are within the scope of the present invention.

EXPRESSION OF DESATURASE GENES

Once the DNA encoding a desaturase polypeptide has been obtained, it is placed in a vector capable of replication in a host cell, or is propagated *in vitro* by means of techniques such as PCR or long PCR. Replicating vectors can include plasmids, phage, viruses, cosmids and the like. Desirable vectors include those useful for mutagenesis of the gene of interest or for expression of the gene of interest in host cells. The technique of long PCR has made *in vitro* propagation of large constructs possible, so that modifications to the gene of interest, such as mutagenesis or addition of expression signals, and propagation of the resulting constructs can occur entirely *in vitro* without the use of a replicating vector or a host cell.

For expression of a desaturase polypeptide, functional transcriptional and translational initiation and termination regions are operably linked to the DNA encoding the desaturase polypeptide. Expression of the polypeptide coding region can take place *in vitro* or in a host cell. Transcriptional and translational initiation and termination regions are derived from a variety of nonexclusive sources, including the DNA to be expressed, genes known or suspected to be capable of expression in the desired system, expression vectors, chemical synthesis, or from an endogenous locus in a host cell.

Expression In Vitr

In vitro expression can be accomplished, for example, by placing the coding region for the desaturase polypeptide in an expression vector designed for *in vitro* use and adding rabbit reticulocyte lysate and cofactors; labeled amino acids can be incorporated if desired. Such *in vitro* expression vectors may provide some or all of the expression signals necessary in the system used. These methods are well known in the art and the components of the system are commercially available. The reaction mixture can then be assayed directly for the polypeptide, for example by determining its activity, or the synthesized polypeptide can be purified and then assayed.

Expression In A Host Cell

Expression in a host cell can be accomplished in a transient or stable fashion. Transient expression can occur from introduced constructs which contain expression signals functional in the host cell, but which constructs do not replicate and rarely integrate in the host cell, or where the host cell is not proliferating. Transient expression also can be accomplished by inducing the activity of a regulatable promoter operably linked to the gene of interest, although such inducible systems frequently exhibit a low basal level of expression. Stable expression can be achieved by introduction of a construct that can integrate into the host genome or that autonomously replicates in the host cell. Stable expression of the gene of interest can be selected for through the use of a selectable marker located on or transfected with the expression construct, followed by selection for cells expressing the marker. When stable expression results from integration, integration of constructs can occur randomly within the host genome or can be targeted through the use of constructs containing regions of homology with the host genome sufficient to target recombination with the host locus. Where constructs are targeted to an endogenous locus, all or some of the transcriptional and translational regulatory regions can be provided by the endogenous locus.

When increased expression of the desaturase polypeptide in the source organism is desired, several methods can be employed. Additional genes encoding the desaturase polypeptide can be introduced into the host organism. Expression from the native desaturase locus also can be increased through homologous recombination, for example by inserting a stronger promoter into the host genome to cause increased expression, by removing destabilizing sequences from either the mRNA or the encoded protein by deleting that information from the host genome, or by adding stabilizing sequences to the mRNA (USPN 4,910,141).

When it is desirable to express more than one different gene, appropriate regulatory regions and expression methods, introduced genes can be propagated in the host cell through use of replicating vectors or by integration into the host genome. Where two or more genes are expressed from separate replicating vectors, it is desirable that each vector has a different means of replication. Each introduced construct, whether integrated or not, should have a different means of selection and should lack homology to the other constructs to maintain stable expression and prevent reassortment of elements among constructs. Judicious choices of regulatory regions, selection means and method of propagation of the introduced construct can be experimentally determined so that all introduced genes are expressed at the necessary levels to provide for synthesis of the desired products.

As an example, where the host cell is a yeast, transcriptional and translational regions functional in yeast cells are provided, particularly from the host species. The transcriptional initiation regulatory regions can be obtained, for example from genes in the glycolytic pathway, such as alcohol dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase (GPD), phosphoglucoisomerase, phosphoglycerate kinase, etc. or regulatable genes such as acid phosphatase, lactase, metallothionein, glucoamylase, etc. Any one of a number of regulatory sequences can be used in a particular situation, depending upon whether constitutive or induced transcription is desired, the particular efficiency of the promoter in conjunction with the open-reading frame of interest, the ability to join a strong promoter with a control region from a

different promoter which allows for inducible transcription, ease f construction, and the like. Of particular interest are promoters which are activated in the presence of galactose. Galactose-inducible promoters (GAL1, GAL7, and GAL10) have been extensively utilized for high level and regulated expression of protein in yeast (Lue *et al.*, *Mol. Cell. Biol.* Vol. 7, p. 3446, 1987; Johnston, *Microbiol. Rev.* Vol. 51, p. 458, 1987). Transcription from the GAL promoters is activated by the GAL4 protein, which binds to the promoter region and activates transcription when galactose is present. In the absence of galactose, the antagonist GAL80 binds to GAL4 and prevents GAL4 from activating transcription. Addition of galactose prevents GAL80 from inhibiting activation by GAL4.

Nucleotide sequences surrounding the translational initiation codon ATG have been found to affect expression in yeast cells. If the desired polypeptide is poorly expressed in yeast, the nucleotide sequences of exogenous genes can be modified to include an efficient yeast translation initiation sequence to obtain optimal gene expression. For expression in *Saccharomyces*, this can be done by site-directed mutagenesis of an inefficiently expressed gene by fusing it in-frame to an endogenous *Saccharomyces* gene, preferably a highly expressed gene, such as the lactase gene.

The termination region can be derived from the 3' region of the gene from which the initiation region was obtained or from a different gene. A large number of termination regions are known to and have been found to be satisfactory in a variety of hosts from the same and different genera and species. The termination region usually is selected more as a matter of convenience rather than because of any particular property. Preferably, the termination region is derived from a yeast gene, particularly *Saccharomyces*, *Schizosaccharomyces*, *Candida* or *Kluyveromyces*. The 3' regions of two mammalian genes, γ interferon and α2 interferon, are also known to function in yeast.

INTRODUCTION OF CONSTRUCTS INTO HOST CELLS

Constructs comprising the gene of interest may be introduced into a host cell by standard techniques. These techniques include transformation, protoplast fusion, lipofection, transfection, transduction, conjugation, infection, bolistic impact, electroporation, microinjection, scraping, or any other method which introduces the gene of interest into the host cell. Methods of transformation which are used include lithium acetate transformation (*Methods in Enzymology*, Vol. 194, p. 186-187, 1991). For convenience, a host cell which has been manipulated by any method to take up a DNA sequence or construct will be referred to as "transformed" or "recombinant" herein.

The subject host will have at least have one copy of the expression construct and may have two or more, depending upon whether the gene is integrated into the genome, amplified, or is present on an extrachromosomal element having multiple copy numbers. Where the subject host is a yeast, four principal types of yeast plasmid vectors can be used: Yeast Integrating plasmids (YIps), Yeast Replicating plasmids (YRps), Yeast Centromere plasmids (YCps), and Yeast Episomal plasmids (YEps). YIps lack a yeast replication origin and must be propagated as integrated elements in the yeast genome. YRps have a chromosomally derived autonomously replicating sequence and are propagated as medium copy number (20 to 40), autonomously replicating, unstably segregating plasmids. YCps have both a replication origin and a centromere sequence and propagate as low copy number (10-20), autonomously replicating, stably segregating plasmids. YEps have an origin of replication from the yeast 2μm plasmid and are propagated as high copy number, autonomously replicating, irregularly segregating plasmids. The presence of the plasmids in yeast can be ensured by maintaining selection for a marker on the plasmid. Of particular interest are the yeast vectors pYES2 (a YEp plasmid available from Invitrogen, confers uracil prototrophy and a GAL1 galactose-inducible promoter for expression), pRS425-pG1 (a YEp plasmid obtained from Dr. T. H. Chang, Ass. Professor of Molecular Genetics, Ohio State University, containing a constitutive GPD promoter and conferring leucine prototrophy), and pYX424 (a YEp plasmid having a constitutive TP1 promoter and conferring

leucine prototrophy; Alber, T. and Kawasaki, G. (1982). *J. Mol. & Appl. Genetics* 1: 419).

The transformed host cell can be identified by selection for a marker contained on the introduced construct. Alternatively, a separate marker construct may be introduced with the desired construct, as many transformation techniques introduce many DNA molecules into host cells. Typically, transformed hosts are selected for their ability to grow on selective media. Selective media may incorporate an antibiotic or lack a factor necessary for growth of the untransformed host, such as a nutrient or growth factor. An introduced marker gene therefor may confer antibiotic resistance, or encode an essential growth factor or enzyme, and permit growth on selective media when expressed in the transformed host. Selection of a transformed host can also occur when the expressed marker protein can be detected, either directly or indirectly. The marker protein may be expressed alone or as a fusion to another protein. The marker protein can be detected by its enzymatic activity; for example β galactosidase can convert the substrate X-gal to a colored product, and luciferase can convert luciferin to a light-emitting product. The marker protein can be detected by its light-producing or modifying characteristics; for example, the green fluorescent protein of *Aequorea victoria* fluoresces when illuminated with blue light. Antibodies can be used to detect the marker protein or a molecular tag on, for example, a protein of interest. Cells expressing the marker protein or tag can be selected, for example, visually, or by techniques such as FACS or panning using antibodies. For selection of yeast transformants, any marker that functions in yeast may be used. Desirably, resistance to kanamycin and the amino glycoside G418 are of interest, as well as ability to grow on media lacking uracil, leucine, lysine or tryptophan.

Of particular interest is the Δ6- and Δ12-desaturase-mediated production of PUFAs in prokaryotic and eukaryotic host cells. Prokaryotic cells of interest include *Eschericia*, *Bacillus*, *Lactobacillus*, *cyanobacteria* and the like. Eukaryotic cells include mammalian cells such as those of lactating animals, avian cells such as of chickens, and other cells amenable to genetic manipulation including insect, fungal, and algae cells. The cells may be

cultured or formed as part or all of a host organism including an animal. Viruses and bacteriophage also may be used with the cells in the production of PUFAs, particularly for gene transfer, cellular targeting and selection. In a preferred embodiment, the host is any microorganism or animal which produces and/or can assimilate exogenously supplied substrate(s) for a Δ6- and/or Δ12-desaturase, and preferably produces large amounts of one or more of the substrates. Examples of host animals include mice, rats, rabbits, chickens, quail, turkeys, bovines, sheep, pigs, goats, yaks, etc., which are amenable to genetic manipulation and cloning for rapid expansion of the transgene expressing population. For animals, the desaturase transgene(s) can be adapted for expression in target organelles, tissues and body fluids through modification of the gene regulatory regions. Of particular interest is the production of PUFAs in the breast milk of the host animal.

Expression In Yeast

Examples of host microorganisms include *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, or other yeast such as *Candida, Kluyveromyces* or other fungi, for example, filamentous fungi such as *Aspergillus, Neurospora, Penicillium*, etc. Desirable characteristics of a host microorganism are, for example, that it is genetically well characterized, can be used for high level expression of the product using ultra-high density fermentation, and is on the GRAS (generally recognized as safe) list since the proposed end product is intended for ingestion by humans. Of particular interest is use of a yeast, more particularly baker's yeast (*S. cerevisiae*), as a cell host in the subject invention. Strains of particular interest are SC334 (Mat α pep4-3 prbl-1122 ura3-52 leu2-3, 112 regl-501 gal1; *Gene* 83:57-64, 1989, Hovland P. *et al.*), YTC34 (α ade2-101 his3Δ200 lys2-801 ura3-52; obtained from Dr. T. H. Chang, Ass. Professor of Molecular Genetics, Ohio State University), YTC41 (a/α ura3-52/ura3=52 lys2-801/lys2-801 ade2-101/ade2-101 trp1-Δ1/trp1-Δ1 his3Δ200/his3Δ200 leu2Δ1/leu2Δ1; obtained from Dr. T. H. Chang, Ass. Professor of Molecular Genetics, Ohio State University), BJ1995 (obtained from the Yeast Genetic

Stock Centre, 1021 Donner Laboratory, Berkeley, CA 94720), INVSC1 (Mat α hiw3Δ1 leu2 trp1-289 ura3-52; obtained from Invitrogen, 1600 Faraday Ave., Carlsbad, CA 92008) and INVSC2 (Mat α his3Δ200 ura3-167; obtained from Invitrogen).

Expression in Avian Species

For producing PUFAs in avian species and cells, such as chickens, turkeys, quail and ducks, gene transfer can be performed by introducing a nucleic acid sequence encoding a Δ6 and/or Δ12-desaturase into the cells following procedures known in the art. If a transgenic animal is desired, pluripotent stem cells of embryos can be provided with a vector carrying a desaturase encoding transgene and developed into adult animal (USPN 5,162,215; Ono *et al.* (1996) *Comparative Biochemistry and Physiology A 113*(3):287-292; WO 9612793; WO 9606160). In most cases, the transgene will be modified to express high levels of the desaturase in order to increase production of PUFAs. The transgene can be modified, for example, by providing transcriptional and/or translational regulatory regions that function in avian cells, such as promoters which direct expression in particular tissues and egg parts such as yolk. The gene regulatory regions can be obtained from a variety of sources, including chicken anemia or avian leukosis viruses or avian genes such as a chicken ovalbumin gene.

Expression in Insect Cells

Production of PUFAs in insect cells can be conducted using baculovirus expression vectors harboring one or more desaturase transgenes. Baculovirus expression vectors are available from several commercial sources such as Clonetech. Methods for producing hybrid and transgenic strains of algae, such as marine algae, which contain and express a desaturase transgene also are provided. For example, transgenic marine algae may be prepared as described in USPN 5,426,040. As with the other expression systems described above, the timing, extent of expression and activity of the desaturase transgene can be

regulated by fitting the polypeptide coding sequence with the appropriate transcriptional and translational regulatory regions selected for a particular use. Of particular interest are promoter regions which can be induced under preselected growth conditions. For example, introduction of temperature sensitive and/or metabolite responsive mutations into the desaturase transgene coding sequences, its regulatory regions, and/or the genome of cells into which the transgene is introduced can be used for this purpose.

The transformed host cell is grown under appropriate conditions adapted for a desired end result. For host cells grown in culture, the conditions are typically optimized to produce the greatest or most economical yield of PUFAs, which relates to the selected desaturase activity. Media conditions which may be optimized include: carbon source, nitrogen source, addition of substrate, final concentration of added substrate, form of substrate added, aerobic or anaerobic growth, growth temperature, inducing agent, induction temperature, growth phase at induction, growth phase at harvest, pH, density, and maintenance of selection. Microorganisms of interest, such as yeast are preferably grown in selected medium. For yeast, complex media such as peptone broth (YPD) or a defined media such as a minimal media (contains amino acids, yeast nitrogen base, and ammonium sulfate, and lacks a component for selection, for example uracil) are preferred. Desirably, substrates to be added are first dissolved in ethanol. Where necessary, expression of the polypeptide of interest may be induced, for example by including or adding galactose to induce expression from a GAL promoter.

Expression In Plants

Production of PUFA's in plants can be conducted using various plant transformation systems such as the use of *Agrobacterium tumefaciens*, plant viruses, particle cell transformation and the like which are disclosed in Applicant's related applications U.S. Application Serial Nos. 08/834,033 and 08/956,985 and continuation-in-part applications filed simultaneously with this application all of which are hereby incorporated by reference.

Expression In An Animal

Expression in cells of a host animal can likewise be accomplished in a transient or stable manner. Transient expression can be accomplished via known methods, for example infection or lipofection, and can be repeated in order to maintain desired expression levels of the introduced construct (*see* Ebert, PCT publication WO 94/05782). Stable expression can be accomplished via integration of a construct into the host genome, resulting in a transgenic animal. The construct can be introduced, for example, by microinjection of the construct into the pronuclei of a fertilized egg, or by transfection, retroviral infection or other techniques whereby the construct is introduced into a cell line which may form or be incorporated into an adult animal (U.S. Patent No. 4,873,191; U.S. Patent No. 5,530,177; U.S. Patent No. 5,565,362; U.S. Patent No. 5,366,894; Willmut *et al* (1997) Nature 385:810). The recombinant eggs or embryos are transferred to a surrogate mother (U.S. Patent No. 4,873,191; U.S. Patent No. 5,530,177; U.S. Patent No. 5,565,362; U.S. Patent No. 5,366,894; Wilmut *et al* (supra)).

After birth, transgenic animals are identified, for example, by the presence of an introduced marker gene, such as for coat color, or by PCR or Southern blotting from a blood, milk or tissue sample to detect the introduced construct, or by an immunological or enzymological assay to detect the expressed protein or the products produced therefrom (U.S. Patent No. 4,873,191; U.S. Patent No. 5,530,177; U.S. Patent No. 5,565,362; U.S. Patent No. 5,366,894; Wilmut *et al* (supra)). The resulting transgenic animals may be entirely transgenic or may be mosaics, having the transgenes in only a subset of their cells. The advent of mammalian cloning, accomplished by fusing a nucleated cell with an enucleated egg, followed by transfer into a surrogate mother, presents the possibility of rapid, large-scale production upon obtaining a "founder" animal or cell comprising the introduced construct; prior to this, it was necessary for the transgene to be present in the germ line of the animal for propagation (Wilmut *et al* (supra)).

Expression in a host animal presents certain efficiencies, particularly where the host is a domesticated animal. For production of PUFAs in a fluid readily obtainable from the host animal, such as milk, the desaturase transgene can be expressed in mammary cells from a female host, and the PUFA content of the host cells altered. The desaturase transgene can be adapted for expression so that it is retained in the mammary cells, or secreted into milk, to form the PUFA reaction products localized to the milk (PCT publication WO 95/24488). Expression can be targeted for expression in mammary tissue using specific regulatory sequences, such as those of bovine α-lactalbumin, α-casein, β-casein, γ-casein, κ-casein, β-lactoglobulin, or whey acidic protein, and may optionally include one or more introns and/or secretory signal sequences (U.S. Patent No. 5,530,177; Rosen, U.S. Patent No. 5,565,362; Clark *et al.*, U.S. Patent No. 5,366,894; Garner *et al.*, PCT publication WO 95/23868). Expression of desaturase transgenes, or antisense desaturase transcripts, adapted in this manner can be used to alter the levels of specific PUFAs, or derivatives thereof, found in the animals milk. Additionally, the desaturase transgene(s) can be expressed either by itself or with other transgenes, in order to produce animal milk containing higher proportions of desired PUFAs or PUFA ratios and concentrations that resemble human breast milk (Prieto *et al.*, PCT publication WO 95/24494).

PURIFICATION OF FATTY ACIDS

The desaturated fatty acids may be found in the host microorganism or animal as free fatty acids or in conjugated forms such as acylglycerols, phospholipids, sulfolipids or glycolipids, and may be extracted from the host cell through a variety of means well-known in the art. Such means may include extraction with organic solvents, sonication, supercritical fluid extraction using for example carbon dioxide, and physical means such as presses, or combinations thereof. Of particular interest is extraction with hexane or methanol and chloroform. Where desirable, the aqueous layer can be acidified to protonate negatively charged moieties and thereby increase partitioning of desired products into the organic layer. After extraction, the organic solvents can be removed by evaporation under a stream of nitrogen. When isolated in

conjugated forms, the products may be enzymatically or chemically cleaved to release the free fatty acid or a less complex conjugate of interest, and can then be subject to further manipulations to produce a desired end product. Desirably, conjugated forms of fatty acids are cleaved with potassium hydroxide.

If further purification is necessary, standard methods can be employed. Such methods may include extraction, treatment with urea, fractional crystallization, HPLC, fractional distillation, silica gel chromatography, high speed centrifugation or distillation, or combinations of these techniques. Protection of reactive groups, such as the acid or alkenyl groups, may be done at any step through known techniques, for example alkylation or iodination. Methods used include methylation of the fatty acids to produce methyl esters. Similarly, protecting groups may be removed at any step. Desirably, purification of fractions containing GLA, SDA, ARA, DHA and EPA may be accomplished by treatment with urea and/or fractional distillation.

USES OF FATTY ACIDS

The fatty acids of the subject invention finds many applications. Probes based on the DNAs of the present invention may find use in methods for isolating related molecules or in methods to detect organisms expressing desaturases. When used as probes, the DNAs or oligonucleotides must be detectable. This is usually accomplished by attaching a label either at an internal site, for example via incorporation of a modified residue, or at the 5' or 3' terminus. Such labels can be directly detectable, can bind to a secondary molecule that is detectably labeled, or can bind to an unlabelled secondary molecule and a detectably labeled tertiary molecule; this process can be extended as long as is practical to achieve a satisfactorily detectable signal without unacceptable levels of background signal. Secondary, tertiary, or bridging systems can include use of antibodies directed against any other molecule, including labels or other antibodies, or can involve any molecules which bind to each other, for example a biotin-streptavidin/avidin system. Detectable labels typically include radioactive isotopes, molecules which chemically or enzymatically produce or alter light, enzymes which produce

detectable reaction products, magnetic molecules, fluorescent m lecules or molecules whose fluorescence or light-emitting characteristics change upon binding. Examples of labelling methods can be found in USPN 5,011,770. Alternatively, the binding of target molecules can be directly detected by measuring the change in heat of solution on binding of probe to target via isothermal titration calorimetry, or by coating the probe or target on a surface and detecting the change in scattering of light from the surface produced by binding of target or probe, respectively, as may be done with the BIAcore system.

PUFAs produced by recombinant means find applications in a wide variety of areas. Supplementation of animals or humans with PUFAs in various forms can result in increased levels not only of the added PUFAs but of their metabolic progeny as well.

NUTRITIONAL COMPOSITIONS

The present invention also includes nutritional compositions. Such compositions, for purposes of the present invention, include any food or preparation for human consumption including for enteral or parenteral consumption, which when taken into the body (a) serve to nourish or build up tissues or supply energy and/or (b) maintain, restore or support adequate nutritional status or metabolic function.

The nutritional composition of the present invention comprises at least one oil or acid produced in accordance with the present invention and may either be in a solid or liquid form. Additionally, the composition may include edible macronutrients, vitamins and minerals in amounts desired for a particular use. The amount of such ingredients will vary depending on whether the composition is intended for use with normal, healthy infants, children or adults having specialized needs such as those which accompany certain metabolic conditions (e.g., metabolic disorders).

Examples of macronutrients which may be added to the composition include but are not limited to edible fats, carbohydrates and proteins. Examples of such edible fats include but are not limited to coconut oil, soy oil, and mono-

and diglycerides. Examples of such carbohydrates include but are not limited to glucose, edible lactose and hydrolyzed search. Additionally, examples of proteins which may be utilized in the nutritional composition of the invention include but are not limited to soy proteins, electrodialysed whey , electrodialysed skim milk, milk whey, or the hydrolysates of these proteins.

With respect to vitamins and minerals, the following may be added to the nutritional compositions of the present invention: calcium, phosphorus, potassium, sodium, chloride, magnesium, manganese, iron, copper, zinc, selenium, iodine, and Vitamins A, E, D, C, and the B complex. Other such vitamins and minerals may also be added.

The components utilized in the nutritional compositions of the present invention will of semi-purified or purified origin. By semi-purified or purified is meant a material which has been prepared by purification of a natural material or by synthesis.

Examples of nutritional compositions of the present invention include but are not limited to infant formulas, dietary supplements, and rehydration compositions. Nutritional compositions of particular interest include but are not limited to those utilized for enteral and parenteral supplementation for infants, specialist infant formulae, supplements for the elderly, and supplements for those with gastrointestinal difficulties and/or malabsorption.

Nutritional Compositions

A typical nutritional composition of the present invention will contain edible macronutrients, vitamins and minerals in amounts desired for a particular use. The amounts of such ingredients will vary depending on whether the formulation is intended for use with normal, healthy individuals temporarily exposed to stress, or to subjects having specialized needs due to certain chronic or acute disease states (e.g., metabolic disorders). It will be understood by persons skilled in the art that the components utilized in a nutritional formulation of the present invention are of semi-purified or purified origin. By semi-purified or purified is meant a material that has been prepared by

purification of a natural material or by synthesis. These techniques are well known in the art (See, e.g., Code of Federal Regulations for Food Ingredients and Food Processing; Recommended Dietary Allowances, 10th Ed., National Academy Press, Washington, D.C., 1989).

In a preferred embodiment, a nutritional formulation of the present invention is an enteral nutritional product, more preferably an adult or child enteral nutritional product. Accordingly in a further aspect of the invention, a nutritional formulation is provided that is suitable for feeding adults or children, who are experiencing stress. The formula comprises, in addition to the PUFAs of the invention; macronutrients, vitamins and minerals in amounts designed to provide the daily nutritional requirements of adults.

The macronutritional components include edible fats, carbohydrates and proteins. Exemplary edible fats are coconut oil, soy oil, and mono- and diglycerides and the PUFA oils of this invention. Exemplary carbohydrates are glucose, edible lactose and hydrolyzed cornstarch. A typical protein source would be soy protein, electrodialysed whey or electrodialysed skim milk or milk whey, or the hydrolysates of these proteins, although other protein sources are also available and may be used. These macronutrients would be added in the form of commonly accepted nutritional compounds in amount equivalent to those present in human milk or an energy basis, i.e., on a per calorie basis.

Methods for formulating liquid and enteral nutritional formulas are well known in the art and are described in detail in the examples.

The enteral formula can be sterilized and subsequently utilized on a ready-to-feed (RTF) basis or stored in a concentrated liquid or a powder. The powder can be prepared by spray drying the enteral formula prepared as indicated above, and the formula can be reconstituted by rehydrating the concentrate. Adult and infant nutritional formulas are well known in the art and commercially available (e.g., Similac®, Ensure®, Jevity® and Alimentum® from Ross Products Division, Abbott Laboratories). An oil or acid of the present invention can be added to any of these formulas in the amounts described below.

The energy density of the nutritional composition when in liquid form, can typically range from about 0.6 to 3.0 Kcal per ml. When in solid or powdered form, the nutritional supplement can contain from about 1.2 to more than 9 Kcals per gm, preferably 3 to 7 Kcals per gm. In general, the osmolality of a liquid product should be less than 700 mOsm and more preferably less than 660 mOsm.

The nutritional formula would typically include vitamins and minerals, in addition to the PUFAs of the invention, in order to help the individual ingest the minimum daily requirements for these substances. In addition to the PUFAs listed above, it may also be desirable to supplement the nutritional composition with zinc, copper, and folic acid in addition to antioxidants. It is believed that these substances will also provide a boost to the stressed immune system and thus will provide further benefits to the individual. The presence of zinc, copper or folic acid is optional and is not required in order to gain the beneficial effects on immune suppression. Likewise a pharmaceutical composition can be supplemented with these same substances as well.

In a more preferred embodiment, the nutritional contains, in addition to the antioxidant system and the PUFA component, a source of carbohydrate wherein at least 5 weight % of said carbohydrate is an indigestible oligosaccharide. In yet a more preferred embodiment, the nutritional composition additionally contains protein, taurine and carnitine.

The PUFAs, or derivatives thereof, made by the disclosed method can be used as dietary substitutes, or supplements, particularly infant formulas, for patients undergoing intravenous feeding or for preventing or treating malnutrition. Typically, human breast milk has a fatty acid profile comprising from about 0.15 % to about 0.36 % as DHA, from about 0.03 % to about 0.13 % as EPA, from about 0.30 % to about 0.88 % as ARA, from about 0.22 % to about 0.67 % as DGLA, and from about 0.27 % to about 1.04 % as GLA. Additionally, the predominant triglyceride in human milk has been reported to be 1,3-di-oleoyl-2-palmitoyl, with 2-palmitoyl glycerides reported as better absorbed than 2-oleoyl or 2-lineoyl glycerides (USPN 4,876,107). Thus, fatty acids such as ARA, DGLA, GLA and/or EPA produced by the invention can be

used to alter the composition of infant formulas to better replicate the PUFA composition of human breast milk. In particular, an oil composition for use in a pharmacologic or food supplement, particularly a breast milk substitute or supplement, will preferably comprise one or more of ARA, DGLA and GLA. More preferably the oil will comprise from about 0.3 to 30% ARA, from about 0.2 to 30% DGLA, and from about 0.2 to about 30% GLA.

In addition to the concentration, the ratios of ARA, DGLA and GLA can be adapted for a particular given end use. When formulated as a breast milk supplement or substitute, an oil composition which contains two or more of ARA, DGLA and GLA will be provided in a ratio of about 1:19:30 to about 6:1:0.2, respectively. For example, the breast milk of animals can vary in ratios of ARA:DGLA:DGL ranging from 1:19:30 to 6:1:0.2, which includes intermediate ratios which are preferably about 1:1:1, 1:2:1, 1:1:4. When produced together in a host cell, adjusting the rate and percent of conversion of a precursor substrate such as GLA and DGLA to ARA can be used to precisely control the PUFA ratios. For example, a 5% to 10% conversion rate of DGLA to ARA can be used to produce an ARA to DGLA ratio of about 1:19, whereas a conversion rate of about 75% to 80% can be used to produce an ARA to DGLA ratio of about 6:1. Therefore, whether in a cell culture system or in a host animal, regulating the timing, extent and specificity of desaturase expression as described can be used to modulate the PUFA levels and ratios. Depending on the expression system used, e.g., cell culture or an animal expressing oil(s) in its milk, the oils also can be isolated and recombined in the desired concentrations and ratios. Amounts of oils providing these ratios of PUFA can be determined following standard protocols. PUFAs, or host cells containing them, also can be used as animal food supplements to alter an animal's tissue or milk fatty acid composition to one more desirable for human or animal consumption.

For dietary supplementation, the purified PUFAs, or derivatives thereof, may be incorporated into cooking oils, fats or margarines formulated so that in normal use the recipient would receive the desired amount. The PUFAs may

also be incorporated into infant formulas, nutritional supplements or other food products, and may find use as anti-inflammatory or cholesterol lowering agents.

Pharmaceutical Compositions

The present invention also encompasses a pharmaceutical composition comprising one or more of the acids and/or resulting oils produced in accordance with the methods described herein. More specifically, such a pharmaceutical composition may comprise one or more of the acids and/or oils as well as a standard, well-known, non-toxic pharmaceutically acceptable carrier, adjuvant or vehicle such as, for example, phosphate buffered saline, water, ethanol, polyols, vegetable oils, a wetting agent or an emulsion such as a water/oil emulsion. The composition may be in either a liquid or solid form. For example, the composition may be in the form of a tablet, capsule, ingestible liquid or powder, injectible, or topical ointment or cream.

Possible routes of administration include, for example, oral, rectal and parenteral. The route of administration will, of course, depend upon the desired effect. For example, if the composition is being utilized to treat rough, dry, or aging skin, to treat injured or burned skin, or to treat skin or hair affected by a disease or condition, it may perhaps be applied topically.

The dosage of the composition to be administered to the patient may be determined by one of ordinary skill in the art and depends upon various factors such as weight of the patient, age of the patient, immune status of the patient, etc.

With respect to form, the composition may be, for example, a solution, a dispersion, a suspension, an emulsion or a sterile powder which is then reconstituted.

Additionally, the composition of the present invention may be utilized for cosmetic purposes. It may be added to pre-existing cosmetic compositions such that a mixture is formed or may be used as a sole composition.

Pharmaceutical compositions may be utilized to administer the PUFA component to an individual. Suitable pharmaceutical compositions may comprise physiologically acceptable sterile aqueous or non-aqueous solutions, dispersions, suspensions or emulsions and sterile powders for reconstitution into sterile solutions or dispersions for ingestion. Examples of suitable aqueous and non-aqueous carriers, diluents, solvents or vehicles include water, ethanol, polyols (propyleneglycol, polyethyleneglycol, glycerol, and the like), suitable mixtures thereof, vegetable oils (such as olive oil) and injectable organic esters such as ethyl oleate. Proper fluidity can be maintained, for example, by the maintenance of the required particle size in the case of dispersions and by the use of surfactants. It may also be desirable to include isotonic agents, for example sugars, sodium chloride and the like. Besides such inert diluents, the composition can also include adjuvants, such as wetting agents, emulsifying and suspending agents, sweetening, flavoring and perfuming agents.

Suspensions, in addition to the active compounds, may contain suspending agents, as for example, ethoxylated isostearyl alcohols, polyoxyethylene sorbitol and sorbitan esters, microcrystalline cellulose, aluminum metahydroxide, bentonite, agar-agar and tragacanth or mixtures of these substances, and the like.

Solid dosage forms such as tablets and capsules can be prepared using techniques well known in the art. For example, PUFAs of the invention can be tableted with conventional tablet bases such as lactose, sucrose, and cornstarch in combination with binders such as acacia, cornstarch or gelatin, disintegrating agents such as potato starch or alginic acid and a lubricant such as stearic acid or magnesium stearate. Capsules can be prepared by incorporating these excipients into a gelatin capsule along with the antioxidants and the PUFA component. The amount of the antioxidants and PUFA component that should be incorporated into the pharmaceutical formulation should fit within the guidelines discussed above.

As used in this application, the term "treat" refers to either preventing, or reducing the incidence of, the undesired occurrence. For example, to treat immune suppression refers to either preventing the occurrence of this

suppression or reducing the amount of such suppression. The terms "patient" and "individual" are being used interchangeably and both refer to an animal. The term "animal" as used in this application refers to any warm-blooded mammal including, but not limited to, dogs, humans, monkeys, and apes. As used in the application the term "about" refers to an amount varying from the stated range or number by a reasonable amount depending upon the context of use. Any numerical number or range specified in the specification should be considered to be modified by the term about.

"Dose" and "serving" are used interchangeably and refer to the amount of the nutritional or pharmaceutical composition ingested by the patient in a single setting and designed to deliver effective amounts of the antioxidants and the structured triglyceride. As will be readily apparent to those skilled in the art, a single dose or serving of the liquid nutritional powder should supply the amount of antioxidants and PUFAs discussed above. The amount of the dose or serving should be a volume that a typical adult can consume in one sitting. This amount can vary widely depending upon the age, weight, sex or medical condition of the patient. However as a general guideline, a single serving or dose of a liquid nutritional produce should be considered as encompassing a volume from 100 to 600 ml, more preferably from 125 to 500 ml and most preferably from 125 to 300 ml.

The PUFAs of the present invention may also be added to food even when supplementation of the diet is not required. For example, the composition may be added to food of any type including but not limited to margarines, modified butters, cheeses, milk, yogurt, chocolate, candy, snacks, salad oils, cooking oils, cooking fats, meats, fish and beverages.

<u>Pharmaceutical Applications</u>

For pharmaceutical use (human or veterinary), the compositions are generally administered orally but can be administered by any route by which they may be successfully absorbed, e.g., parenterally (i.e. subcutaneously, intramuscularly or intravenously), rectally or vaginally or topically, for example, as a skin ointment or lotion. The PUFAs of the present invention may

be administered alone or in combination with a pharmaceutically acceptable carrier or excipient. Where available, gelatin capsules are the preferred form of oral administration. Dietary supplementation as set forth above also can provide an oral route of administration. The unsaturated acids of the present invention may be administered in conjugated forms, or as salts, esters, amides or prodrugs of the fatty acids. Any pharmaceutically acceptable salt is encompassed by the present invention; especially preferred are the sodium, potassium or lithium salts. Also encompassed are the N-alkylpolyhydroxamine salts, such as N-methyl glucamine, found in PCT publication WO 96/33155. The preferred esters are the ethyl esters. As solid salts, the PUFAs also can be administered in tablet form. For intravenous administration, the PUFAs or derivatives thereof may be incorporated into commercial formulations such as Intralipids. The typical normal adult plasma fatty acid profile comprises 6.64 to 9.46% of ARA, 1.45 to 3.11% of DGLA, and 0.02 to 0.08% of GLA. These PUFAs or their metabolic precursors can be administered, either alone or in mixtures with other PUFAs, to achieve a normal fatty acid profile in a patient. Where desired, the individual components of formulations may be individually provided in kit form, for single or multiple use. A typical dosage of a particular fatty acid is from 0.1 mg to 20 g, or even 100 g daily, and is preferably from 10 mg to 1, 2, 5 or 10 g daily as required, or molar equivalent amounts of derivative forms thereof. Parenteral nutrition compositions comprising from about 2 to about 30 weight percent fatty acids calculated as triglycerides are encompassed by the present invention; preferred is a composition having from about 1 to about 25 weight percent of the total PUFA composition as GLA (USPN 5,196,198). Other vitamins, and particularly fat-soluble vitamins such as vitamin A, D, E and L-carnitine can optionally be included. Where desired, a preservative such as α tocopherol may be added, typically at about 0.1% by weight.

Suitable pharmaceutical compositions may comprise physiologically acceptable sterile aqueous or non-aqueous solutions, dispersions, suspensions or emulsions and sterile powders for reconstitution into sterile injectible solutions or dispersions. Examples of suitable aqueous and non-aqeuous carriers,

diluents, solvents or vehicles include water, ethanol, polyols (propylleneglyol, polyethylenegycol, glycerol, and the like), suitable mixtures thereof, vegetable oils (such as olive oil) and injectable organic esters such as ehyl oleate. Proper fluidity can be maintained, for example, by the maintenance of the required particle size in the case of dispersions and by the use of surfactants. It may also be desirable to include isotonic agents, for example sugars, sodium chloride and the like. Besides such inert diluents, the composition can also include adjuvants, such as wetting agents, emulsifying and suspending agents, sweetening, flavoring and perfuming agents.

Suspensions in addition to the active compounds, may contain suspending agents, as for example, ethoxylated isostearyl alcohols, polyoxyethylene sorbitol and sorbitan esters, microcrystalline cellulose, aluminum metahydroxide, bentonite, agar-agar and tragacanth, or mixtures of these substances and the like.

An especially preferred pharmaceutical composition contains diacetyltartaric acid esters of mono- and diglycerides dissolved in an aqueous medium or solvent. Diacetyltartaric acid esters of mono- and diglycerides have an HLB value of about 9-12 and are significantly more hydrophilic than existing antimicrobial lipids that have HLB values of 2-4. Those existing hydrophobic lipids cannot be formulated into aqueous compositions. As disclosed herein, those lipids can now be solubilized into aqueous media in combination with diacetyltartaric acid esters of mono-and diglycerides. In accordance with this embodiment, diacetyltartaric acid esters of mono- and diglycerides (e.g., DATEM-C12:0) is melted with other active antimicrobial lipids (e.g., 18:2 and 12:0 monoglycerides) and mixed to obtain a homogeneous mixture. Homogeneity allows for increased antimicrobial activity. The mixture can be completely dispersed in water. This is not possible without the addition of diacetyltartaric acid esters of mono- and diglycerides and premixing with other monoglycerides prior to introduction into water. The aqueous composition can then be admixed under sterile conditions with physiologically acceptable diluents, preservatives, buffers or propellants as may be required to form a spray or inhalant.

The present invention also encompasses the treatment of numerous disorders with fatty acids. Supplementation with PUFAs of the present invention can be used to treat restenosis after angioplasty. Symptoms of inflammation, rheumatoid arthritis, and asthma and psoriasis can be treated with the PUFAs of the present invention. Evidence indicates that PUFAs may be involved in calcium metabolism, suggesting that PUFAs of the present invention may be used in the treatment or prevention of osteoporosis and of kidney or urinary tract stones.

The PUFAs of the present invention can be used in the treatment of cancer. Malignant cells have been shown to have altered fatty acid compositions; addition of fatty acids has been shown to slow their growth and cause cell death, and to increase their susceptibility to chemotherapeutic agents. GLA has been shown to cause reexpression on cancer cells of the E-cadherin cellular adhesion molecules, loss of which is associated with aggressive metastasis. Clinical testing of intravenous administration of the water soluble lithium salt of GLA to pancreatic cancer patients produced statistically significant increases in their survival. PUFA supplementation may also be useful for treating cachexia associated with cancer.

The PUFAs of the present invention can also be used to treat diabetes (USPN 4,826,877; Horrobin *et al.*, Am. J. Clin. Nutr. Vol. 57 (Suppl.), 732S-737S). Altered fatty acid metabolism and composition has been demonstrated in diabetic animals. These alterations have been suggested to be involved in some of the long-term complications resulting from diabetes, including retinopathy, neuropathy, nephropathy and reproductive system damage. Primrose oil, which contains GLA, has been shown to prevent and reverse diabetic nerve damage.

The PUFAs of the present invention can be used to treat eczema, reduce blood pressure and improve math scores. Essential fatty acid deficiency has been suggested as being involved in eczema, and studies have shown beneficial effects on eczema from treatment with GLA. GLA has also been shown to reduce increases in blood pressure associated with stress, and to improve performance on arithmetic tests. GLA and DGLA have been shown to inhibit

platelet aggregation, cause vasodilation, lower cholesterol levels and inhibit proliferation of vessel wall smooth muscle and fibrous tissue (Brenner *et al.*, Adv. Exp. Med. Biol. Vol. 83, p. 85-101, 1976). Administration of GLA or DGLA, alone or in combination with EPA, has been shown to reduce or prevent gastro-intestinal bleeding and other side effects caused by non-steroidal anti-inflammatory drugs (USPN 4,666,701). GLA and DGLA have also been shown to prevent or treat endometriosis and premenstrual syndrome (USPN 4,758,592) and to treat myalgic encephalomyelitis and chronic fatigue after viral infections (USPN 5,116,871).

Further uses of the PUFAs of this invention include use in treatment of AIDS, multiple schlerosis, acute respiratory syndrome, hypertension and inflammatory skin disorders. The PUFAs of the inventions also can be used for formulas for general health as well as for geriatric treatments.

Veterinary Applications

It should be noted that the above-described pharmaceutical and nutritional compositions may be utilized in connection with animals, as well as humans, as animals experience many of the same needs and conditions as human. For example, the oil or acids of the present invention may be utilized in animal feed supplements.

The following examples are presented by way of illustration, not of limitation.

Examples

Example 1	Construction of a cDNA Library from *Mortierella alpina*
Example 2	Isolation of a Δ6-desaturase Nucleotide Sequence from *Mortierella alpina*
Example 3	Identification of Δ6-desaturases Homologous to the *Mortierella alpina* Δ6-desaturase
Example 4	Isolation of a Δ12-desaturase Nucleotide Sequence from *Mortierella Alpina*

Example 5	Expression of *M. alpina* Desaturase Clones in Baker's Yeast
Example 6	Initial Optimization of Culture Conditions
Example 7	Distribution of PUFAs in Yeast Lipid Fractions
Example 8	Further Culture Optimization and Coexpression of Δ6 and Δ12-desaturases
Example 9	Identification of Homologues to *M. alpina* Δ5 and Δ6 desaturases
Example 10	Identification of *M. alpina* Δ5 and Δ6 homologues in other PUFA-producing organisms
Example 11	Identification of *M. alpina* Δ5 and Δ6 homologues in other PUFA-producing organisms
Example 12	Human Desaturase Gene Sequences
Example 13	Nutritional Compositions

Example 1

Construction of a cDNA Library from *Mortierella alpina*

Total RNA was isolated from a 3 day old PUFA-producing culture of *Mortierella alpina* using the protocol of Hoge *et al.* (1982) *Experimental Mycology* 6:225-232. The RNA was used to prepare double-stranded cDNA using BRL's lambda-ZipLox system following the manufactures instructions. Several size fractions of the *M. alpina* cDNA were packaged separately to yield libraries with different average-sized inserts. A "full-length" library contains approximately 3 x 10^6 clones with an average insert size of 1.77 kb. The "sequencing-grade" library contains approximately 6 x 10^5 clones with an average insert size of 1.1 kb.

Example 2

Isolation of a Δ6-desaturase Nucleotide Sequence from *Mortierella Alpina*

A nucleic acid sequence from a partial cDNA clone, Ma524, encoding a Δ6 fatty acid desaturase from *Mortierella alpina* was obtained by random sequencing of clones from the *M. alpina* cDNA sequencing grade library described in Example 1. cDNA-containing plasmids were excised as follows:

Five µl of phage were combined with 100 µl of *E. coli* DH10B(ZIP) grown in ECLB plus 10 µg/ml kanamycin, 0.2% maltose, and 10 mM $MgSO_4$ and incubated at 37 degrees for 15 minutes. 0.9 ml SOC was added and 100 µl of the bacteria immediately plated on each of 10 ECLB + 50 µg Pen plates. No 45 minute recovery time was needed. The plates were incubated overnight at 37°. Colonies were picked into ECLB + 50 µg Pen media for overnight cultures to be used for making glycerol stocks and miniprep DNA. An aliquot of the culture used for the miniprep is stored as a glycerol stock. Plating on ECLB + 50 µg Pen/ml resulted in more colonies and a greater proportion of colonies containing inserts than plating on 100 µg/ml Pen.

Random colonies were picked and plasmid DNA purified using Qiagen miniprep kits. DNA sequence was obtained from the 5' end of the cDNA insert and compared to the National Center for Biotechnology Information (NCBI) nonredundant database using the BLASTX algorithm. Ma524 was identified as a putative desaturase based on DNA sequence homology to previously identified desaturases.

A full-length cDNA clone was isolated from the *M. alpina* full-length library and designed pCGN5532. The cDNA is contained as a 1617 bp insert in the vector pZL1 (BRL) and, beginning with the first ATG, contains an open reading frame encoding 457 amino acids. The three conserved "histidine boxes" known to be conserved among membrane-bound deaturases (Okuley, et al. (1994) *The Plant Cell* 6:147-158) were found to be present at amino acid positions 172-176, 209-213, and 395-399 (see Figure 3). As with other

membrane-bound Δ6-desaturases the final HXXHH histidine box motif was found to be QXXHH. The amino acid sequence of Ma524 was found to display significant homology to a portion of a *Caenorhabditis elegans* cosmid, WO6D2.4, a cytochrome b5/desaturase fusion protein from sunflower, and the *Synechocystis* and *Spirulina* Δ 6-desaturases. In addition, Ma524 was shown to have homology to the borage Δ6-desaturase amino sequence (PCT publication W) 96/21022). Ma524 thus appears to encode a Δ6-desaturase that is related to the borage and algal Δ6-desaturases. The peptide sequences are shown as SEQ ID NO:5 - SEQ ID NO:11.

The amino terminus of the encoded protein was found to exhibit significant homology to cytochrome b5 proteins. The *Mortierella* cDNA clone appears to represent a fusion between a cytochrome b5 and a fatty acid desaturase. Since cytochrome b5 is believed to function as the electron donor for membrane-bound desaturase enzymes, it is possible that the N-terminal cytochrome b5 domain of this desaturase protein is involved in its function. This may be advantageous when expressing the desaturase in heterologous systems for PUFA production. However, it should be noted that, although the amino acid sequences of Ma524 and the borage Δ6 were found to contain regions of homology, the base compositions of the cDNAs were shown to be significantly different. For example, the borage cDNA was shown to have an overall base composition of 60 % A/T, with some regions exceeding 70 %, while Ma524 was shown to have an average of 44 % A/T base composition, with no regions exceeding 60 %. This may have implications for expressing the cDNAs in microorganisms or animals which favor different base compositions. It is known that poor expression of recombinant genes can occur when the host prefers a base composition different from that of the introduced gene. Mechanisms for such poor expression include decreased stability, cryptic splice sites, and/or translatability of the mRNA and the like.

Example 3

Identification of Δ6-desaturases Homologous to the *Mortierella alpina* Δ6-desaturase

Nucleic acid sequences that encode putative Δ6-desaturases were identified through a BLASTX search of the Expressed Sequence Tag ("EST") databases through NCBI using the Ma524 amino acid sequence. Several sequences showed significant homology. In particular, the deduced amino acid sequence of two *Arabidopsis thaliana* sequences, (accession numbers F13728 and T42806) showed homology to two different regions of the deduced amino acid sequence of Ma524. The following PCR primers were designed: ATTS4723-FOR (complementary to F13728) SEQ ID NO:13
5' CUACUACUACUAGGAGTCCTCTACGGTGTTTTG and
T42806-REV (complementary to T42806) SEQ ID NO:14
5' CAUCAUCAUCAUATGATGCTCAAGCTGAAACTG. Five μg of total RNA isolated from developing siliques of *Arabidopsis thaliana* was reverse transcribed using BRL Superscript RTase and the primer TSyn (5'-CCAAGCTTCTGCAGGAGCTCTTTTTTTTTTTTTTT-3') and is shown as SEQ ID NO:12. PCR was carried out in a 50 ul volume containing: template derived from 25 ng total RNA, 2 pM each primer, 200 μM each deoxyribonucleotide triphosphate, 60 mM Tris-Cl, pH 8.5, 15 mM $(NH_4)_2SO_4$, 2 mM $MgCl_2$, 0.2 U Taq Polymerase. Thermocycler conditions were as follows: 94 degrees for 30 sec., 50 degrees for 30 sec., 72 degrees for 30 sec. PCR was continued for 35 cycles followed by an additional extension at 72 degrees for 7 minutes. PCR resulted in a fragment of approximately ~750 base pairs which was subcloned, named 12-5, and sequenced. Each end of this fragment was formed to correspond to the *Arabidopsis* ESTs from which the PCR primers were designed. The putative amino acid sequence of 12-5 was compared to that of Ma524, and ESTs from human (W28140), mouse (W53753), and *C. elegans* (R05219) (see Figure 4). Homology patterns with the *Mortierella* Δ6- desaturase indicate that these sequences represent putative

desaturase polypeptides. Based on this experiment approach, it is likely that the full-length genes can be cloned using probes based on the EST sequences. Following the cloning, the genes can then be placed into expression vectors, expressed in host cells, and their specific Δ6- or other desaturase activity can be determined as described below.

Example 4

Isolation of a Δ12-desaturase Nucleotide Sequence from *Mortierella alpina*

Based on the fatty acids it accumulates, it seemed probable that *Mortierella alpina* has an ω6 type desaturase. The ω6-desaturase is responsible for the production of linoleic acid (18:2) from oleic acid (18:1). Linoleic acid (18:2) is a substrate for a Δ6-desaturase. This experiment was designed to determine if *Mortierella alpina* has a Δ12-desaturase polypeptide, and if so, to identify the corresponding nucleotide sequence.

A random colony from the *M. alpina* sequencing grade library, Ma648, was sequenced and identified as a putative desaturase based on DNA sequence homology to previously identified desaturases, as described for Ma524 (*see* Example 2). The nucleotide sequence is shown in SEQ ID NO:13. The peptide sequence is shown in SEQ ID NO:4. The deduced amino acid sequence from the 5' end of the Ma648 cDNA displays significant homology to soybean microsomal ω6 (Δ12) desaturase (accession #L43921) as well as castor bean oleate 12-hydroxylase (accession #U22378). In addition, homology was observed when compared to a variety of other ω6 (Δ12) and ω3 (Δ15) fatty acid desaturase sequences.

Example 5

Expression of *M. alpina* Desaturase Clones in Baker's Yeast

Yeast Transformation

Lithium acetate transformation of yeast was performed according to standard protocols (*Methods in Enzymology*, Vol. 194, p. 186-187, 1991). Briefly, yeast were grown in YPD at 30°C. Cells were spun down, resuspended in TE, spun down again, resuspended in TE containing 100 mM lithium acetate, spun down again, and resuspended in TE/lithium acetate. The resuspended yeast were incubated at 30°C for 60 minutes with shaking. Carrier DNA was added, and the yeast were aliquoted into tubes. Transforming DNA was added, and the tubes were incubated for 30 min. at 30°C. PEG solution (35% (w/v) PEG 4000, 100 mM lithium acetate, TE pH7.5) was added followed by a 50 min. incubation at 30°C. A 5 min. heat shock at 42°C was performed, the cells were pelleted, washed with TE, pelleted again and resuspended in TE. The resuspended cells were then plated on selective media.

Desaturase Expression in Transformed Yeast

cDNA clones from *Mortierella alpina* were screened for desaturase activity in baker's yeast. A canola Δ15-desaturase (obtained by PCR using 1st strand cDNA from *Brassica napus* cultivar 212/86 seeds using primers based on the published sequence (Arondel *et al. Science* 258:1353-1355)) was used as a positive control. The Δ15-desaturase gene and the gene from cDNA clones Ma524 and Ma648 were put in the expression vector pYES2 (Invitrogen), resulting in plasmids pCGR-2, pCGR-5 and pCGR-7, respectively. These plasmids were transfected into *S. cerevisiae* yeast strain 334 and expressed after induction with galactose and in the presence of substrates that allowed detection of specific desaturase activity. The control strain was *S. cerevisiae* strain 334 containing the unaltered pYES2 vector. The substrates used, the products produced and the indicated desaturase activity were: DGLA (conversion to ARA would indicate Δ5-desaturase activity), linoleic acid (conversion to GLA

would indicate Δ6-desaturase activity; conversion to ALA would indicate Δ15-desaturase activity), oleic acid (an endogenous substrate made by *S. cerevisiae*, conversion to linoleic acid would indicate Δ12-desaturase activity, which *S. cerevisiae* lacks), or ARA (conversion to EPA would indicate Δ17-desaturase activity).

Cultures were grown for 48-52 hours at 15°C in the presence of a particular substrate. Lipid fractions were extracted for analysis as follows: Cells were pelleted by centrifugation, washed once with sterile ddH_20, and repelleted. Pellets were vortexed with methanol; chloroform was added along with tritridecanoin (as an internal standard). The mixtures were incubated for at least one hour at room temperature or at 4°C overnight. The chloroform layer was extracted and filtered through a Whatman filter with one gram of anhydrous sodium sulfate to remove particulates and residual water. The organic solvents were evaporated at 40°C under a stream of nitrogen. The extracted lipids were then derivatized to fatty acid methyl esters (FAME) for gas chromatography analysis (GC) by adding 2 ml of 0.5 N potassium hydroxide in methanol to a closed tube. The samples were heated to 95°C to 100°C for 30 minutes and cooled to room temperature. Approximately 2 ml of 14 % boron trifluoride in methanol was added and the heating repeated. After the extracted lipid mixture cooled, 2 ml of water and 1 ml of hexane were added to extract the FAME for analysis by GC. The percent conversion was calculated by dividing the product produced by the sum of (the product produced and the substrate added) and then multiplying by 100. To calculate the oleic acid percent conversion, as no substrate was added, the total linoleic acid produced was divided by the sum of oleic acid and linoleic acid produced, then multiplying by 100. The desaturase activity results are provided in Table 1 below.

Table 1

M. alpina Desaturase Expression in Baker's Yeast

CLONE	ENZYME ACTIVITY	% CONVERSION OF SUBSTRATE
pCGR-2	Δ6	0 (18:2 to 18:3w6)
(canola Δ15	Δ15	16.3 (18:2 to 18:3w3)
desaturase)	Δ5	2.0 (20:3 to 20:4w6)
	Δ17	2.8 (20:4 to 20:5w3)
	Δ12	1.8 (18:1 to 18:2w6)
pCGR-5	Δ6	6.0
(M. alpina	Δ15	0
Ma524	Δ5	2.1
	Δ17	0
	Δ12	3.3
pCGR-7	Δ6	0
(M. alpina	Δ15	3.8
Ma648	Δ5	2.2
	Δ17	0
	Δ12	63.4

The Δ15-desaturase control clone exhibited 16.3% conversion of the substrate. The pCGR-5 clone expressing the Ma524 cDNA showed 6% conversion of the substrate to GLA, indicating that the gene encodes a Δ6-desaturase. The pCGR-7 clone expressing the Ma648 cDNA converted 63.4% conversion of the substrate to LA, indicating that the gene encodes a Δ12-desaturase. The background (non-specific conversion of substrate) was between 0-3% in these cases. We also found substrate inhibition of the activity by using different concentrations of the substrate. When substrate was added to 100 μM, the percent conversion to product dropped compared to when substrate was added to 25 μM (see below). Additionally, by varying the substrate concentration between 5 μM and 200 μM, conversion ratios were found to range between about

5% to about 75% greater. These data show that desaturases with different substrate specificities can be expressed in a heterologous system and used to produce poly-unsaturated long chain fatty acids.

Table 2 represents fatty acids of interest as a percent of the total lipid extracted from the yeast host *S. cerevisiae* 334 with the indicated plasmid. No glucose was present in the growth media. Affinity gas chromatography was used to separate the respective lipids. GC/MS was employed to verify the identity of the product(s). The expected product for the *B. napus* Δ15-desaturase, α-linolenic acid, was detected when its substrate, linoleic acid, was added exogenously to the induced yeast culture. This finding demonstrates that yeast expression of a desaturase gene can produce functional enzyme and detectable amounts of product under the current growth conditions. Both exogenously added substrates were taken up by yeast, although slightly less of the longer chain PUFA, dihomo-γ-linolenic acid (20:3), was incorporated into yeast than linoleic acid (18:2) when either was added in free form to the induced yeast cultures. γ-linolenic acid was detected when linoleic acid was present during induction and expression of *S. cerevisiae* 334 (pCGR-5). The presence of this PUFA demonstrates Δ6-desaturase activity from pCGR-5 (MA524). Linoleic acid, identified in the extracted lipids from expression of *S. cerevisiae* 334 (pCGR-7), classifies the cDNA MA648 from *M. alpina* as the Δ12-desaturase.

Table 2

Fatty Acid as a Percentage of Total Lipid Extracted from Yeast

Plasmid in Yeast (enzyme)	18:2 Incorporated	α-18:3 Produced	γ-18:3 Produced	20:3 Incorporated	20:4 Produced	18:1* Present	18:2 Produced
pYES2 (control)	66.9	0	0	58.4	0	4	0
pCGR-2 (Δ15)	60.1	5.7	0	50.4	0	0.7	0
pCGR-5 (Δ6)	62.4	0	4.0	49.9	0	2.4	0
pCGR-7 (Δ12)	65.6	0	0	45.7	0	7.1	12.2

100 μM substrate added

* 18:1 is an endogenous fatty acid in yeast

Key To Tables
18:1=oleic acid
18:2=linoleic acid
α-18:3=α-linolenic acid
γ-18:3=γ-linolenic acid
18:4=stearidonic acid
20:3=dihomo-γ-linolenic acid
20:4=arachidonic acid

Example 6

Optimization of Culture Conditions

Table 3A shows the effect of exogenous free fatty acid substrate concentration on yeast uptake and conversion to fatty acid product as a percentage of the total yeast lipid extracted. In all instances, low amounts of exogenous substrate (1-10 μM) resulted in low fatty acid substrate uptake and product formation. Between 25 and 50 μM concentration of free fatty acid in the growth and induction media gave the highest percentage of fatty acid product formed, while the 100 μM concentration and subsequent high uptake into yeast appeared to decrease or inhibit the desaturase activity. The amount of fatty acid substrate for yeast expressing Δ12-desaturase was similar under the same growth conditions, since the substrate, oleic acid, is an endogenous yeast fatty acid. The use of α-linolenic acid as an additional substrate for pCGR-5 (Δ6) produced the expected product, stearidonic acid (Table 3A). The feedback inhibition of high fatty acid substrate concentration was well illustrated when the percent conversion rates of the respective fatty acid substrates to their respective products were compared in Table 3B. In all cases, 100 μM substrate concentration in the growth media decreased the percent conversion to product. The uptake of α-linolenic was comparable to other PUFAs added in free form, while the Δ6-desaturase percent conversion, 3.8-17.5%, to the product stearidonic acid was the lowest of all the substrates examined (Table 3B). The effect of media, such as YPD (rich media) versus minimal media with glucose on the conversion rate of Δ12-desaturase was dramatic. Not only did the conversion rate for oleic to linoleic acid drop, (Table 3B) but the percent of linoleic acid formed also decreased by 11% when rich media was used for growth and induction of yeast desaturase Δ12 expression (Table 3A). The effect of media composition was also evident when glucose was present in the growth media for Δ6-desaturase, since the percent of substrate uptake was decreased at 25 μM (Table 3A). However, the conversion rate remained the

same and percent product formed decreased for Δ6-desaturase for in the presence of glucose.

Table 3A

Effect of Added Substrate on the Percentage of Incorporated Substrate and Product Formed in Yeast Extracts

Plasmid in Yeast	pCGR-2 (Δ15)	PcGR-5 (Δ6)	pCGR-5 (Δ6)	pCGR-7 (Δ12)
Substrate/product	18:2 /α-18:3	18:2/γ-18:3	α-18:3/18:4	18:1*/18:2
1 μM sub.	ND	0.9/0.7	ND	ND
10μM sub.	ND	4.2/2.4	10.4/2.2	ND
25 μM sub.	ND	11/3.7	18.2/2.7	ND
25 μM◊ sub.	36.6/7.2◊	25.1/10.3◊	ND	6.6/15.8◊
50 μM sub.	53.1/6.5◊	ND	36.2/3	10.8/13^{+}
100 μM sub.	60.1/5.7◊	62.4/4◊	47.7/1.9	10/24.8

Table 3B

Effect of Substrate Concentration in Media on the Percent Conversion of Fatty Acid Substrate to Product in Yeast Extracts

Plasmid in Yeast	pCGR-2 (Δ15)	pCGR-5 (Δ6)	pCGR-5 (Δ6)	pCGR-7 (Δ12)
substrate→product	18:2 →α-18:3	18:2→γ18:3	α-18:3→18:4	18:1*→18:2
1 μM sub.	ND	43.8	ND	ND
10 μM sub.	ND	36.4	17.5	ND
25 μM sub.	ND	25.2	12.9	ND
25 μM◊ sub.	16.4◊	29.1◊	ND	70.5◊
50 μM sub.	10.9◊	ND	7.7	54.6⁺
100 μM sub.	8.7◊	6◊	3.8	71.3

◊ no glucose in media

⁺ Yeast peptone broth (YPD)

* 18:1 is an endogenous yeast lipid

sub. is substrate concentration

ND (not done)

Table 4 shows the amount of fatty acid produced by a recombinant desaturase from induced yeast cultures when different amounts of free fatty acid substrate were used. Fatty acid weight was determined since the total amount of lipid varied dramatically when the growth conditions were changed, such as the presence of glucose in the yeast growth and induction media. To better determine the conditions when the recombinant desaturase would produce the most PUFA product, the quantity of individual fatty acids were examined. The absence of glucose dramatically reduced by three fold the amount of linoleic acid produced by recombinant Δ12-desaturase. For the Δ12-desaturase the amount of total yeast lipid was decreased by almost half in the absence of glucose. Conversely, the presence of glucose in the yeast growth media for Δ6-desaturase drops the γ-linolenic acid produced by almost half, while the total amount of yeast lipid produced was not changed by the presence/absence of

glucose. This points to a possible role for glucose as a modulator of Δ6-desaturase activity.

Table 4

Fatty Acid Produced in µg from Yeast Extracts

Plasmid in Yeast (enzyme)	pCGR-5 (Δ6)	pCGR-5 (Δ6)	pCGR-7 (Δ12)
product	Υ-18:3	18:4	18:2*
1 µM sub.	1.9	ND	ND
10 µM sub.	5.3	4.4	ND
25 µM sub.	10.3	8.7	115.7
25 µM ◊ sub.	29.6	ND	39 ◊

◊ no glucose in media
sub. is substrate concentration
ND (not done)
*18:1, the substrate, is an endogenous yeast lipid

Example 7

Distribution of PUFAs in Yeast Lipid Fractions

Table 5 illustrates the uptake of free fatty acids and their new products formed in yeast lipids as distributed in the major lipid fractions. A total lipid extract was prepared as described above. The lipid extract was separated on TLC plates, and the fractions were identified by comparison to standards. The bands were collected by scraping, and internal standards were added. The fractions were then saponified and methylated as above, and subjected to gas chromatography. The gas chromatograph calculated the amount of fatty acid by comparison to a standard. The phospholipid fraction contained the highest amount of substrate and product PUFAs for Δ6-desaturase activity. It would appear that the substrates are accessible in the phospholipid form to the desaturases.

Table 5

Fatty Acid Distribution in Various Yeast Lipid Fractions in μg

Fatty acid fraction	Phospholipid	Diglyceride	Free Fatty Acid	Triglyceride	Cholesterol Ester
SC (pCGR-5) substrate 18:2	166.6	6.2	15	18.2	15.6
SC (pCGR-5) product γ-18:3	61.7	1.6	4.2	5.9	1.2

SC = *S. cerevisiae* (plasmid)

Example 8

Further Culture Optimization and Coexpression of Δ6 and Δ12-desaturases

This experiment was designed to evaluate the growth and induction conditions for optimal activities of desaturases in *Saccharomyces cerevisiae*. A *Saccharomyces cerevisiae* strain (SC334) capable of producing γ-linolenic acid (GLA) was developed, to assess the feasibility of production of PUFA in yeast. The genes for Δ6 and Δ12-desaturases from *M. alpina* were coexpressed in SC334. Expression of Δ12-desaturase converted oleic acid (present in yeast) to linoleic acid. The linoleic acid was used as a substrate by the Δ6-desaturase to produce GLA. The quantity of GLA produced ranged between 5-8% of the total fatty acids produced in SC334 cultures and the conversion rate of linoleic acid to γ-linolenic acid ranged between 30% to 50%. The induction temperature was optimized, and the effect of changing host strain and upstream promoter sequences on expression of Δ6 and Δ12 (MA 524 and MA 648 respectively) desaturase genes was also determined.

Plasmid Construction

The cloning of pCGR5 as well as pCGR7 has been discussed above. To construct pCGR9a and pCGR9b, the Δ6 and Δ12-desaturase genes were amplified using the following sets of primers. The primers pRDS1 and 3 had Xhol site and primers pRDS2 and 4 had Xbal site (indicated in bold). These primer sequences are presented as SEQ ID NO:15-18.

I. Δ6-desaturase amplification primers

a. pRDS1 TAC CAA **CTC GAG** AAA ATG GCT GCT GCT CCC AGT GTG AGG

b. pRDS2 AAC TGA **TCT AGA** TTA CTG CGC CTT ACC CAT CTT GGA GGC

II. Δ12-desaturase amplification primers

a. pRDS3 TAC CAA **CTC GAG** AAA ATG GCA CCT CCC AAC ACT ATC GAT

b. pRDS4 AAC TGA **TCT AGA** TTA CTT CTT GAA AAA GAC CAC GTC TCC

The pCGR5 and pCGR7 constructs were used as template DNA for amplification of Δ6 and Δ12-desaturase genes, respectively. The amplified products were digested with Xbal and Xhol to create "sticky ends". The PCR amplified Δ6-desaturase with Xhol-Xbal ends as cloned into pCGR7, which was also cut with Xho-l-Xbal. This procedure placed the Δ6-desaturase behind the Δ12-desaturase, under the control of an inducible promoter GAL1. This construct was designated pCGR9a. Similarly, to construct pCGR9b, the Δ12-desaturase with XhoI-XbaI ends was cloned in the XhoI-XbaI sites of pCGR5. In pCGR9b the Δ12-desaturase was behind the Δ6-desaturase gene, away from the GAL promoter.

To construct pCGR10, the vector pRS425, which contains the constitutive Glyceraldehyde 3-Phosphate Dehydrogenase (GPD) promoter, was digested with BamHl and pCGR5 was digested with BamHl-Xhol to release the

Δ6-desaturase gene. This Δ6-desaturase fragment and BamHl cut pRS425 were filled using Klenow Polymerase to create blunt ends and ligated, resulting in pCGR10a and pCGR10b containing the Δ6-desaturase gene in the sense and antisense orientation, respectively. To construct pCGR11 and pCGR12, the Δ6 and Δ12-desaturase genes were isolated from pCGR5 and pCGR7, respectively, using an EcoRl-XhoI double digest. The EcoRl-Xhol fragments of Δ6 and Δ12-desaturases were cloned into the pYX242 vector digested with EcoRl-Xhol. The pYX242 vector has the promoter of TPl (a yeast housekeeping gene), which allows constitutive expression.

Yeast Transformation and Expression

Different combinations of pCGR5, pCGR7, pCGR9a, pCGR9b, pCGR10a, pCGR11 and pCGR12 were introduced into various host strains of *Saccharomyces cerevisiae*. Transformation was done using PEG/LiAc protocol (Methods in Enzymology Vol. 194 (1991): 186-187). Transformants were selected by plating on synthetic media lacking the appropriate amino acid. The pCGR5, pCGR7, pCGR9a and pCGR9b can be selected on media lacking uracil. The pCGR10, pCGR11 and pCGR12 constructs can be selected on media lacking leucine. Growth of cultures and fatty acid analysis was performed as in Example 5 above.

Production of GLA

Production of GLA requires the expression of two enzymes (the Δ6 and Δ12-desaturases), which are absent in yeast. To express these enzymes at optimum levels the following constructs or combinations of constructs, were introduced into various host strains:

1) pCGR9a/SC334
2) pCGR9b/SC334
3) pCGR10a and pCGR7/SC334
4) pCGR11 and pCGR7/SC334
5) pCGR12 and pCGR5/SC334

6) pCGR10a and pCGR7/DBY746

7) pCGR10a and pCGR7/DBY746

The pCGR9a construct has both the Δ6 and Δ12-desaturase genes under the control of an inducible GAL promoter. The SC334 host cells transformed with this construct did not show any GLA accumulation in total fatty acids (Fig. 6A and B, lane 1). However, when the Δ6 and Δ12-desaturase genes were individually controlled by the GAL promoter, the control constructs were able to express Δ6- and Δ12-desaturase, as evidenced by the conversion of their respective substrates to products. The Δ12-desaturase gene in pCGR9a was expressed as evidenced by the conversion of 18:1ω9 to 18:2ω6 in pCGR9a/SC334, while the Δ6-desaturase gene was not expressed/active, because the 18:2ω6 was not being converted to 18:3ω6 (Fig. 6A and B, lane 1).

The pCGR9b construct also had both the Δ6 and Δ12-desaturase genes under the control of the GAL promoter but in an inverse order compared to pCGR9a. In this case, very little GLA (<1%) was seen in pCGR9b/SC334 cultures. The expression of Δ12-desaturase was also very low, as evidenced by the low percentage of 18:2ω6 in the total fatty acids (Fig. 6A and B, lane 1).

To test if expressing both enzymes under the control of independent promoters would increase GLA production, the Δ6-desaturase gene was cloned into the pRS425 vector. The construct of pCGR10a has the Δ6-desaturase in the correct orientation, under control of constitutive GPD promoter. The pCGR10b has the Δ6-desaturase gene in the inverse orientation, and serves as the negative control. The pCGR10a/SC334 cells produced significantly higher levels of GLA (5% of the total fatty acids, Fig. 6, lane 3), compared to pCGR9a. Both the Δ6 and Δ12-desaturase genes were expressed at high level because the conversion of 18:1ω9→ 18:2ω6 was 65%, while the conversion of 18:2ω6 → 18:3ω6 (Δ6-desaturase) was 30% (Fig. 6, lane 3). As expected, the negative control pCGR10b/SC334 did not show any GLA.

To further optimize GLA production, the Δ6 and Δ12 genes were introduced into the pYX242 vector, creating pCGR11 and pCGR12

respectively. The pYX242 vector allows for constitutive expression by the TP1 promoter (Alber, T. and Kawasaki, G. (1982). *J. Mol. & Appl. Genetics* 1: 419). The introduction of pCGR11 and pCGR7 in SC334 resulted in approximately 8% of GLA in total fatty acids of SC334. The rate of conversion of 18:1ω9→ 18:2ω6 and 18:2ω6 → 18:3ω6 was approximately 50% and 44% respectively (Fig. 6A and B, lane 4). The presence of pCGR12 and pCGR5 in SC334 resulted in 6.6% GLA in total fatty acids with a conversion rate of approximately 50% for both 18:1ω9 to 18:2ω6 and 18:2ω6 to 18:3ω6, respectively (Fig. 6A and B, lane 5). Thus although the quantity of GLA in total fatty acids was higher in the pCGR11/pCGR7 combination of constructs, the conversion rates of substrate to product were better for the pCGR12/pCGR5 combination.

To determine if changing host strain would increase GLA production, pCGR10a and pCGR7 were introduced into the host strain BJ1995 and DBY746 (obtained from the Yeast Genetic Stock Centre, 1021 Donner Laboratory, Berkeley, CA 94720. The genotype of strain DBY746 is Matα, his3-Δ1, leu2-3, leu2-112, ura3-32, trp1-289, gal). The results are shown in Fig. 7. Changing host strain to BJ1995 did not improve the GLA production, because the quantity of GLA was only 1.31% of total fatty acids and the conversion rate of 18:1ω9 → 18:2ω6 was approximately 17% in BJ1995. No GLA was observed in DBY746 and the conversion of 18:1ω9 → 18:2ω6 was very low (<1% in control) suggesting that a cofactor required for the expression of Δ12-desaturase might be missing in DB746 (Fig. 7, lane 2).

To determine the effect of temperature on GLA production, SC334 cultures containing pCGR10a and pCGR7 were grown at 15°C and 30°C. Higher levels of GLA were found in cultures grown and induced at 15°C than those in cultures grown at 30°C (4.23% vs. 1.68%). This was due to a lower conversion rate of 18:2ω6 → 18:3ω6 at 30°C (11.6% vs. 29% in 15°C) cultures, despite a higher conversion of 18:1ω9 → 18:2ω6 (65% vs. 60% at 30°C (Fig. 8). These results suggest that Δ12- and Δ6-desaturases may have different optimal expression temperatures.

Of the various parameters examined in this study, temperature of growth, yeast host strain and media components had the most significant impact on the expression of desaturase, while timing of substrate addition and concentration of inducer did not significantly affect desaturase expression.

These data show that two DNAs encoding desaturases that can convert LA to GLA or oleic acid to LA can be isolated from *Mortierella alpina* and can be expressed, either individually or in combination, in a heterologous system and used to produce poly-unsaturated long chain fatty acids. Exemplified is the production of GLA from oleic acid by expression of Δ12- and Δ6-desaturases in yeast.

Example 9

Identification of Homologues to *M. alpina* Δ5 and Δ6 desaturases

A nucleic acid sequence that encodes a putative Δ5 desaturase was identified through a TBLASTN search of the expressed sequence tag databases through NCBI using amino acids 100-446 of Ma29 as a query. The truncated portion of the Ma29 sequence was used to avoid picking up homologies based on the cytochrome b5 portion at the N-terminus of the desaturase. The deduced amino acid sequence of an est from *Dictyostelium discoideum* (accession # C25549) shows very significant homology to Ma29 and lesser, but still significant homology to Ma524. The DNA sequence is presented as SEQ ID NO:19. The amino acid sequence is presented as SEQ ID NO:20.

Example 10

Identification of *M. alpina* Δ5 and Δ6 homologues in other PUFA-producing organisms

To look for desaturases involved in PUFA production, a cDNA library was constructed from total RNA isolated from *Phaeodactylum tricornutum*. A plasmid-based cDNA librariy was constructed in pSPORT1 (GIBCO-BRL)

following manufacturer's instructions using a commercially available kit (GIBCO-BRL). Random cDNA clones were sequenced and nucleic acid sequences that encode putative Δ5 or Δ6 desaturases were identified through BLAST search of the databases and comparison to Ma29 and Ma524 sequences.

One clone was identified from the *Phaeodactylum* library with homology to Ma29 and Ma524; it is called 144-011-B12. The DNA sequence is presented as SEQ ID NO:21. The amino acid sequence is presented as SEQ ID NO:22.

Example 11

Identification of *M. alpina* Δ5 and Δ6 homologues in other PUFA-producing organisms

To look for desaturases involved in PUFA production, a cDNA library was constructed from total RNA isolated from *Schizochytrium* species. A plasmid-based cDNA library was constructed in pSPORT1 (GIBCO-BRL) following manufacturer's instructions using a commercially available kit (GIBCO-BRL). Random cDNA clones were sequenced and nucleic acid sequences that encode putative Δ5 or Δ6 desaturases were identified through BLAST search of the databases and comparison to Ma29 and Ma524 sequences.

One clone was identified from the *Schizochytrium* library with homology to Ma29 and Ma524; it is called 81-23-C7. This clone contains a ~1 kb insert. Partial sequence was obtained from each end of the clone using the universal forward and reverse sequencing primers. The DNA sequence from the forward primer is presented as SEQ ID NO:23. The peptide sequence is presented as SEQ ID NO:24. The DNA sequence from the reverse primer is presented as SEQ ID NO:25. The amino acid sequence from the reverse primer is presented as SEQ ID NO:26.

Example 12

Human Desaturase Gene Sequences

Human desaturase gene sequences potentially involved in long chain polyunsaturated fatty acid biosynthesis were isolated based on homology between the human cDNA sequences and *Mortierella alpina* desaturase gene sequences. The three conserved "histidine boxes" known to be conserved among membrane-bound desaturases were found. As with some other membrane-bound desaturases the final HXXHH histidine box motif was found to be QXXHH. The amino acid sequence of the putative human desaturases exhibited homology to *M. alpina* Δ5, Δ6, Δ9, and Δ12 desaturases.

The *M. alpina* Δ5 desaturase and Δ6 desaturase cDNA sequences were used to search the LifeSeq database of Incyte Pharmaceuticals, Inc., Palo Alto, California 94304. The Δ5 desaturase sequence was divided into fragments; 1) amino acid no. 1-150, 2) amino acid no. 151-300, and 3) amino acid no. 301-446. The Δ6 desaturase sequence was divided into three fragments; 1) amino acid no. 1-150, 2) amino acid no. 151-300, and 3) amino acid no. 301-457. These polypeptide fragments were searched against the database using the "tblastn" algorithm. This alogarithm compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands).

The polypeptide fragments 2 and 3 of *M. alpina* Δ5 and Δ6 have homologies with the CloneID sequences as outlined in Table 6. The CloneID represents an individual sequence from the Incyte LifeSeq database. After the "tblastn" results have been reviewed, Clone Information was searched with the default settings of Stringency of >=50, and Productscore <=100 for different CloneID numbers. The Clone Information Results displayed the information including the ClusterID, CloneID, Library, HitID, Hit Description. When selected, the ClusterID number displayed the clone information of all the clones that belong in that ClusterID. The Assemble command assembles all of the CloneID which comprise the ClusterID. The following default settings were

used for GCG (Genetics Computer Group, University of Wisconsin Biotechnology Center, Madison, Wisconsin 53705) Assembly:

Word Size:	7
Minimum Overlap:	14
Stringency:	0.8
Minimum Identity:	14
Maximum Gap:	10
Gap Weight:	8
Length Weight:	2

GCG Assembly Results displayed the contigs generated on the basis of sequence information within the CloneID. A contig is an alignment of DNA sequences based on areas of homology among these sequences. A new sequence (consensus sequence) was generated based on the aligned DNA sequences within a contig. The contig containing the CloneID was identified, and the ambiguous sites of the consensus sequence was edited based on the alignment of the CloneIDs (see SEQ ID NO:27 - SEQ ID NO:32) to generate the best possible sequence. The procedure was repeated for all six CloneID listed in Table 6. This produced five unique contigs. The edited consensus sequences of the 5 contigs were imported into the Sequencher software program (Gene Codes Corporation, Ann Arbor, Michigan 48 105). These consensus sequences were assembled. The contig 2511785 overlaps with contig 3506132, and this new contig was called 2535 (SEQ ID NO:33). The contigs from the Sequencher program were copied into the Sequence Analysis software package of GCG.

Each contig was translated in all six reading frames into protein sequences. The *M. alpina* Δ5 (MA29) and Δ6 (MA524) sequences were compared with each of the translated contigs using the FastA search (a Pearson

and Lipman search for similarity between a query sequence and a group of sequences of the same type (nucleic acid or protein)). Homology among these sequences suggest the open reading frames of each contig. The homology among the *M. alpina* Δ5 and Δ6 to contigs 2535 and 3854933 were utilized to create the final contig called 253538a. Figure 13 is the FastA match of the final contig 253538a and MA29, and Figure 14 is the FastA match of the final contig 253538a and MA524. The DNA sequences for the various contigs are presented in SEQ ID NO:27 -SEQ ID NO:33 The various peptide sequences are shown in SEQ ID NO:34 - SEQ ID NO: 40.

Although the open reading frame was generated by merging the two contigs, the contig 2535 shows that there is a unique sequence in the beginning of this contig which does not match with the contig 3854933. Therefore, it is possible that these contigs were generated from independent desaturase like human genes.

The contig 253538a contains an open reading frame encoding 432 amino acids. It starts with Gln (CAG) and ends with the stop codon (TGA). The contig 253538a aligns with both *M. alpina* Δ5 and Δ6 sequences, suggesting that it could be either of the desaturases, as well as other known desaturases which share homology with each other. The individual contigs listed in Table 18, as well as the intermediate contig 2535 and the final contig 253538a can be utilized to isolate the complete genes for human desaturases.

Uses of the human desaturases

These human sequences can be express in yeast and plants utilizing the procedures described in the preceding examples. For expression in mammalian cells transgenic animals, these genes may provide superior codon bias.

In addition, these sequences can be used to isolate related desaturase genes from other organisms.

Table 6

Sections f the Desaturases	Clone ID fr m LifeSeq Database	Keyw rd

151-300 Δ5	3808675	fatty acid desaturase
301-446 Δ5	354535	Δ6
151-300 Δ6	3448789	Δ6
151-300 Δ6	1362863	Δ6
151-300 Δ6	2394760	Δ6
301-457 Δ6	3350263	Δ6

Example 13

I. INFANT FORMULATIONS

A. Isomil® Soy Formula with Iron.

Usage: As a beverage for infants, children and adults with an allergy or sensitivity to cow's milk. A feeding for patients with disorders for which lactose should be avoided: lactase deficiency, lactose intolerance and galactosemia.

Features:

- Soy protein isolate to avoid symptoms of cow's-milk-protein allergy or sensitivity
- Lactose-free formulation to avoid lactose-associated diarrhea
- Low osmolaity (240 mOsm/kg water) to reduce risk of osmotic diarrhea.
- Dual carbohydrates (corn syrup and sucrose) designed to enhance carbohydrate absorption and reduce the risk of exceeding the absorptive capacity of the damaged gut.
- 1.8 mg of Iron (as ferrous sulfate) per 100 Calories to help prevent iron deficiency.
- Recommended levels of vitamins and minerals.
- Vegetable oils to provide recommended levels of essential fatty acids.
- Milk-white color, milk-like consistency and pleasant aroma.

Ingredients: (Pareve, ®) 85% water, 4.9% corn syrup, 2.6% sugar (sucrose), 2.1% soy oil, 1.9% soy protein isolate, 1.4% coconut oil, 0.15% calcium citrate, 0.11 % calcium phosphate tribasic, potassium citrate, potassium phosphate monobasic, potassium chloride, mono- and disglycerides, soy lecithin, carrageenan, ascorbic acid, L-methionine, magnesium chloride, potassium phosphate dibasic, sodium chloride, choline chloride, taurine, ferrous sulfate, m-inositol, alpha-tocopheryl acetate, zinc sulfate, L-carnitine, niacinamide, calcium pantothenate, cupric sulfate, vitamin A palmitate, thiamine chloride hydrochloride, riboflavin, pyridoxine hydrochloride, folic acid, manganese sulfate, potassium iodide, phylloquinone, biotin, sodium selenite, vitamin D_3 and cyanocobalamin.

B. Isomil® DF Soy Formula For Diarrhea.

Usage: As a short-term feeding for the dietary management of diarrhea in infants and toddlers.

Features:

- First infant formula to contain added dietary fiber from soy fiber specifically for diarrhea management.
- Clinically shown to reduce the duration of loose, watery stools during mild to severe diarrhea in infants.
- Nutritionally complete to meet the nutritional needs of the infant.
- Soy protein isolate with added L-methionine meets or exceeds an infant's requirement for all essential amino acids.
- Lactose-free formulation to avoid lactose-associated diarrhea.
- Low osmolality (240 mOsm/kg water) to reduce the risk of osmotic diarrhea.
- Dual carbohydrates (corn syrup and sucrose) designed to enhance carbohydrate absorption and reduce the risk of exceeding the absorptive capacity of the damaged gut.

- Meets or exceeds the vitamin and mineral levels recommended by the Committee on Nutrition of the American Academy of Pediatrics and required by the Infant Formula Act.

- 1.8 mg of iron (as ferrous sulfate) per 100 Calories to help prevent iron deficiency.

- Vegetable oils to provide recommended levels of essential fatty acids.

Ingredients: (Pareve, ®) 86% water, 4.8% corn syrup, 2.5% sugar (sucrose), 2.1% soy oil, 2.0% soy protein isolate, 1.4% coconut oil, 0.77% soy fiber, 0.12% calcium citrate, 0.11 % calcium phosphate tribasic, 0.10% potassium citrate, potassium chloride, potassium phosphate monobasic, mono- and disglycerides, soy lecithin, carrageenan, magnesium chloride, ascorbic acid, L-methionine, potassium phosphate dibasic, sodium chloride, choline chloride, taurine, ferrous sulfate, m-inositol, alpha-tocopheryl acetate, zinc sulfate, L-carnitine, niacinamide, calcium pantothenate, cupric sulfate, vitamin A palmitate, thiamine chloride hydrochloride, riboflavin, pyridoxine hydrochloride, folic acid, manganese sulfate, potassium iodide, phylloquinone, biotin, sodium selenite, vitamin D_3 and cyanocobalamin.

C. Isomil® SF Sucrose-Free Soy Formula With Iron.

Usage: As a beverage for infants, children and adults with an allergy or sensitivity to cow's-milk protein or an intolerance to sucrose. A feeding for patients with disorders for which lactose and sucrose should be avoided.

Features:

- Soy protein isolate to avoid symptoms of cow's-milk-protein allergy or sensitivity.

- Lactose-free formulation to avoid lactose-associated diarrhea (carbohydrate source is Polycose® Glucose Polymers).

- Sucrose free for the patient who cannot tolerate sucrose.

- Low osmolality (180 mOsm/kg water) to reduce risk of osmotic diarrhea.
- 1.8 mg of iron (as ferrous sulfate) per 100 Calories to help prevent iron deficiency.
- Recommended levels of vitamins and minerals.
- Vegetable oils to provide recommended levels of essential fatty acids.
- Milk-white color, milk-like consistency and pleasant aroma.

Ingredients: (Pareve, ®) 75% water, 11.8% hydrolized cornstarch, 4.1% soy oil, 4.1% soy protein isolate, 2.8% coconut oil, 1.0% modified cornstarch, 0.38% calcium phosphate tribasic, 0.17% potassium citrate, 0.13% potassium chloride, mono- and disglycerides, soy lecithin, magnesium chloride, abscorbic acid, L-methionine, calcium carbonate, sodium chloride, choline chloride, carrageenan, taurine, ferrous sulfate, m-inositol, alpha-tocopheryl acetate, zinc sulfate, L-carnitine, niacinamide, calcium pantothenate, cupric sulfate, vitamin A palmitate, thiamine chloride hydrochloride, riboflavin, pyridoxine hydrochloride, folic acid, manganese sulfate, potassium iodide, phylloquinone, biotin, sodium selenite, vitamin D_3 and cyanocobalamin.

D. Isomil® 20 Soy Formula With Iron Ready To Feed, 20 Cal/fl oz.

Usage: When a soy feeding is desired.

Ingredients: (Pareve, ®) 85% water, 4.9% corn syrup, 2.6% sugar (sucrose), 2.1% soy oil, 1.9% soy protein isolate, 1.4% coconut oil, 0.15% calcium citrate, 0.11% calcium phosphate tribasic, potassium citrate, potassium phosphate monobasic, potassium chloride, mono- and disglycerides, soy lecithin, carrageenan, abscorbic acid, L-methionine, magnesium chloride, potassium phosphate dibasic, sodium chloride, choline chloride, taurine, ferrous sulfate, m-inositol, alpha-tocopheryl acetate, zinc sulfate, L-carnitine, niacinamide, calcium pantothenate, cupric sulfate, vitamin A palmitate, thiamine chloride hydrochloride, riboflavin, pyridoxine hydrochloride, folic

acid, manganese sulfate, potassium iodide, phylloquinone, biotin, sodium selenite, vitamin D_3 and cyanocobalamin.

E. Similac® Infant Formula

Usage: When an infant formula is needed: if the decision is made to discontinue breastfeeding before age 1 year, if a supplement to breastfeeding is needed or as a routine feeding if breastfeeding is not adopted.

Features:

- Protein of appropriate quality and quantity for good growth; heat-denatured, which reduces the risk of milk-associated enteric blood loss.
- Fat from a blend of vegetable oils (doubly homogenized), providing essential linoleic acid that is easily absorbed.
- Carbohydrate as lactose in proportion similar to that of human milk.
- Low renal solute load to minimize stress on developing organs.
- Powder, Concentrated Liquid and Ready To Feed forms.

Ingredients: (®-D) Water, nonfat milk, lactose, soy oil, coconut oil, mono- and diglycerides, soy lecithin, abscorbic acid, carrageenan, choline chloride, taurine, m-inositol, alpha-tocopheryl acetate, zinc sulfate, niacinamid, ferrous sulfate, calcium pantothenate, cupric sulfate, vitamin A palmitate, thiamine chloride hydrochloride, riboflavin, pyridoxine hydrochloride, folic acid, manganese sulfate, phylloquinone, biotin, sodium selenite, vitamin D_3 and cyanocobalamin.

F. Similac® NeoCare Premature Infant Formula With Iron

Usage: For premature infants' special nutritional needs after hospital discharge. Similac NeoCare is a nutritionally complete formula developed to provide premature infants with extra calories, protein, vitamins and minerals needed to promote catch-up growth and support development.

Features:

- Reduces the need for caloric and vitamin supplementation. More calories (22 Cal/fl oz) then standard term formulas (20 Cal/fl oz).
- Highly absorbed fat blend, with medium-chain triglycerides (MCT oil) to help meet the special digestive needs of premature infants.
- Higher levels of protein, vitamins and minerals per 100 Calories to extend the nutritional support initiated in-hospital.
- More calcium and phosphorus for improved bone mineralization.

Ingredients: ®-D Corn syrup solids, nonfat milk, lactose, whey protein concentrate, soy oil, high-oleic safflower oil, fractionated coconut oil (medium-chain triglycerides), coconut oil, potassium citrate, calcium phosphate tribasic, calcium carbonate, ascorbic acid, magnesium chloride, potassium chloride, sodium chloride, taurine, ferrous sulfate, m-inositol, choline chloride, ascorbyl palmitate, L-carnitine, alpha-tocopheryl acetate, zinc sulfate, niacinamide, mixed tocopherols, sodium citrate, calcium pantothenate, cupric sulfate, thiamine chloride hydrochloride, vitamin A palmitate, beta carotene, riboflavin, pyridoxine hydrochloride, folic acid, manganese sulfate, phylloquinone, biotin, sodium selenite, vitamin D_3 and cyanocobalamin.

G. Similac Natural Care Low-Iron Human Milk Fortifier Ready To Use, 24 Cal/fl oz.

Usage: Designed to be mixed with human milk or to be fed alternatively with human milk to low-birth-weight infants.

Ingredients: ®-D Water, nonfat milk, hydrolyzed cornstarch, lactose, fractionated coconut oil (medium-chain triglycerides), whey protein concentrate, soil oil, coconut oil, calcium phosphate tribasic, potassium citrate, magnesium chloride, sodium citrate, ascorbic acid, calcium carbonate, mono- and diglycerides, soy lecithin, carrageenan, choline chloride, m-inositol, taurine, niacinamide, L-carnitine, alpha tocopheryl acetate, zinc sulfate, potassium chloride, calcium pantothenate, ferrous sulfate, cupric sulfate, riboflavin, vitamin A palmitate, thiamine chloride hydrochloride, pyridoxine

hydrochloride, biotin, folic acid, manganese sulfate, phylloquinone, vitamin D_3, sodium selenite and cyanocobalamin.

Various PUFAs of this invention can be substituted and/or added to the infant formulae described above and to other infant formulae known to those in the art..

II. NUTRITIONAL FORMULATIONS

A. ENSURE®

Usage: ENSURE is a low-residue liquid food designed primarily as an oral nutritional supplement to be used with or between meals or, in appropriate amounts, as a meal replacement. ENSURE is lactose- and gluten-free, and is suitable for use in modified diets, including low-cholesterol diets. Although it is primarily an oral supplement, it can be fed by tube.

Patient Conditions:

- For patients on modified diets
- For elderly patients at nutrition risk
- For patients with involuntary weight loss
- For patients recovering from illness or surgery
- For patients who need a low-residue diet

Ingredients:

®-D Water, Sugar (Sucrose), Maltodextrin (Corn), Calcium and Sodium Caseinates, High-Oleic Safflower Oil, Soy Protein Isolate, Soy Oil, Canola Oil, Potassium Citrate, Calcium Phosphate Tribasic, Sodium Citrate, Magnesium Chloride, Magnesium Phosphate Dibasic, Artificial Flavor, Sodium Chloride, Soy Lecithin, Choline Chloride, Ascorbic Acid, Carrageenan, Zinc Sulfate, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Gellan Gum, Niacinamide, Calcium Pantothenate, Manganese Sulfate, Cupric Sulfate, Vitamin A Palmitate, Thiamine Chloride Hydrochloride, Pyridoxine Hydrochloride, Riboflavin, Folic Acid, Sodium Molybdate, Chromium Chloride, Biotin, Potassium Iodide, Sodium Selenate.

B. ENSURE® BARS

Usage: ENSURE BARS are complete, balanced nutrition for supplemental use between or with meals. They provide a delicious, nutrient-rich alternative to other snacks. ENSURE BARS contain <1 g lactose/bar, and Chocolate Fudge Brownie flavor is gluten-free. (Honey Graham Crunch flavor contains gluten.)

Patient Conditions:

- For patients who need extra calories, protein, vitamins and minerals
- Especially useful for people who do not take in enough calories and nutrients
- For people who have the ability to chew and swallow
- Not to be used by anyone with a peanut allergy or any type of allergy to nuts.

Ingredients:

Honey Graham Crunch -- High-Fructose Corn Syrup, Soy Protein Isolate, Brown Sugar, Honey, Maltodextrin (Corn), Crisp Rice (Milled Rice, Sugar [Sucrose], Salt [Sodium Chloride] and Malt), Oat Bran, Partially Hydrogenated Cottonseed and Soy Oils, Soy Polysaccharide, Glycerine, Whey Protein Concentrate, Polydextrose, Fructose, Calcium Caseinate, Cocoa Powder, Artificial Flafors, Canola Oil, High-Oleic Safflower Oil, Nonfat Dry Milk, Whey Powder, Soy Lecithin and Corn Oil. Manufactured in a facility that processes nuts.

Vitamins and Minerals:

Calcium Phosphate Tribasic, Potassium Phosphate Dibasic, Magnesium Oxide, Salt (Sodium Chloride), Potassium Chloride, Ascorbic Acid, Ferric Orthophosphate, Alpha-Tocopheryl Acetate, Niacinamide, Zinc Oxide, Calcium Pantothenate, Copper Gluconate, Manganese Sulfate, Riboflavin, Beta-Carotene, Pyridoxine Hydrochloride, Thiamine Mononitrate, Folic Acid, Biotin,

Chromium Chloride, Potassium Iodide, Sodium Selenate, Sodium Molybdate, Phylloquinone, Vitamin D_3 and Cyanocobalamin.

Protein:

Honey Graham Crunch - The protein source is a blend of soy protein isolate and milk proteins.

Soy protein isolate	74%
Milk proteins	26%

Fat:

Honey Graham Crunch - The fat source is a blend of partially hydrogenated cottonseed and soybean, canola, high oleic safflower, and corn oils, and soy lecithin.

Partially hydrogenated cottonseed and soybean oil	76%
Canola oil	8%
High-oleic safflower oil	8%
Corn oil	4%
Soy lecithin	4%

Carbohydrate:

Honey Graham Crunch - The carbohydrate source is a combination of high-fructose corn syrup, brown sugar, maltodextrin, honey, crisp rice, glycerine, soy polysaccharide, and oat bran.

High-fructose corn syrup	24%
Brown sugar	21%
Maltodextrin	12%
Honey	11%
Crisp rice	9%
Glycerine	9%
Soy polysaccharide	7%
Oat bran	7%\

C. ENSURE® HIGH PROTEIN

Usage: ENSURE HIGH PROTEIN is a concentrated, high-protein liquid food designed for people who require additional calories, protein, vitamins, and minerals in their diets. It can be used as an oral nutritional supplement with or between meals or, in appropriate amounts, as a meal replacement. ENSURE HIGH PROTEIN is lactose- and gluten-free, and is suitable for use by people recovering from general surgery or hip fractures and by patients at risk for pressure ulcers.

Patient Conditions

- For patients who require additional calories, protein, vitamins, and minerals, such as patients recovering from general surgery or hip fractures, patients at risk for pressure ulcers, and patients on low-cholesterol diets

Features-

- Low in saturated fat
- Contains 6 g of total fat and < 5 mg of cholesterol per serving
- Rich, creamy taste
- Excellent source of protein, calcium, and other essential vitamins and minerals
- For low-cholesterol diets
- Lactose-free, easily digested

Ingredients:

Vanilla Supreme: -®-D Water, Sugar (Sucrose), Maltodextrin (Corn), Calcium and Sodium Caseinates, High-Oleic Safflower Oil, Soy Protein Isolate, Soy Oil, Canola Oil, Potassium Citrate, Calcium Phosphate Tribasic, Sodium Citrate, Magnesium Chloride, Magnesium Phosphate Dibasic, Artificial Flavor, Sodium Chloride, Soy Lecithin, Choline Chloride, Ascorbic Acid, Carrageenan, Zinc Sulfate, Ferrous Suffate, Alpha-Tocopheryl Acetate, Gellan Gum, Niacinamide, Calcium Pantothenate, Manganese Sulfate, Cupric Sulfate, Vitamin A Palmitate, Thiamine Chloride Hydrochloride, Pyridoxine Hydrochloride,

Riboflavin, Folio Acid, Sodium Motybdate, Chromium Chloride, Biotin, Potassium Iodide, Sodium Selenate, Phylloquinone, Vitamin D.3 and Cyanocobalarnin.

Protein:

The protein source is a blend of two high-biologic-value proteins: casein and soy.

Sodium and calcium caseinates	85%
Soy protein isolate	15%

Fat:

The fat source is a blend of three oils: high-oleic safflower, canola, and soy.

High-oleic safflower oil	40%
Canola oil	30%
Soy oil	30%

The level of fat in ENSURE HIGH PROTEIN meets American Heart Association (AHA) guidelines. The 6 grams of fat in ENSURE HIGH PROTEIN represent 24% of the total calories, with 2.6% of the fat being from saturated fatty acids and 7.9% from polyunsaturated fatty acids. These values are within the AHA guidelines of $\leq$ 30% of total calories from fat, < 1 0% of the calories from saturated fatty acids, and $\leq$ 1 0% of total calories from polyunsaturated fatty acids.

Carbohydrate:

ENSURE HIGH PROTEIN contains a combination of maltodextrin and sucrose. The mild sweetness and flavor variety (vanilla supreme, chocolate royal, wild berry, and banana), plus VARI-FLAVORSO® Flavor Pacs in pecan, cherry, strawberry, lemon, and orange, help to prevent flavor fatigue and aid in patient compliance.

Vanilla and other nonchocolate flavors

Sucrose	60%

Maltodextrin	40%
Chocolate	
Sucrose	70%
Maltodextrin	30%

D. ENSURE ® LIGHT

Usage: ENSURE LIGHT is a low-fat liquid food designed for use as an oral nutritional supplement with or between meals. ENSURE LIGHT is lactose- and gluten-free, and is suitable for use in modified diets, including low-cholesterol diets.

Patient Conditions:

- For normal-weight or overweight patients who need extra nutrition in a supplement that contains 50% less fat and 20% fewer calories than ENSURE
- For healthy adults who don't eat right and need extra nutrition

Features:

- Low in fat and saturated fat
- Contains 3 g of total fat per serving and < 5 mg cholesterol
- Rich, creamy taste
- Excellent source of calcium and other essential vitamins and minerals
- For low-cholesterol diets
- Lactose-free, easily digested

Ingredients:

French Vanilla: ®-D Water, Maltodextrin (Corn), Sugar (Sucrose), Calcium Caseinate, High-Oleic Safflower Oil, Canola Oil, Magnesium Chloride, Sodium Citrate, Potassium Citrate, Potassium Phosphate Dibasic, Magnesium Phosphate Dibasic, Natural and Artificial Flavor, Calcium Phosphate Tribasic, Cellulose Gel, Choline Chloride, Soy Lecithin, Carrageenan, Salt (Sodium Chloride),

Ascorbic Acid, Cellulose Gum, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Zinc Sulfate, Niacinamide, Manganese Sulfate, Calcium Pantothenate, Cupric Sulfate, Thiamine Chloride Hydrochloride, Vitamin A Palmitate, Pyridoxine Hydrochloride, Riboflavin, Chromium Chloride, Folic Acid, Sodium Molybdate, Biotin, Potassium Iodide, Sodium Selenate, Phylloquinone, Vitamin D_3 and Cyanocobalamin.

Protein:

The protein source is calcium caseinate.

Calcium caseinate	100%

Fat

The fat source is a blend of two oils: high-oleic safflower and canola.

High-oleic safflower oil	70%
Canola oil	30%

The level of fat in ENSURE LIGHT meets American Heart Association (AHA) guidelines. The 3 grams of fat in ENSURE LIGHT represent 13.5% of the total calories, with 1.4% of the fat being from saturated fatty acids and 2.6% from polyunsaturated fatty acids. These values are within the AHA guidelines of $\leq$ 30% of total calories from fat, < 1 0% of the calories from saturated fatty acids, and $\leq$ 1 0% of total calories from polyunsaturated fatty acids.

Carbohydrate

ENSURE LIGHT contains a combination of maltodextrin and sucrose. The chocolate flavor contains corn syrup as well. The mild sweetness and flavor variety (French vanilla, chocolate supreme, strawberry swirl), plus VARI-FLAVORS® Flavor Pacs in pecan, cherry, strawberry, lemon, and orange, help to prevent flavor fatigue and aid in patient compliance.

Vanilla and other nonchocolate flavors

Sucrose	51%
Maltodextrin	49%

Chocolate

Sucrose	47.0%
Corn Syrup	26.5%
Maltodextrin	26.5%

Vitamins and Minerals

An 8-fl-oz serving of ENSURE LIGHT provides at least 25% of the RDIs for 24 key vitamins and minerals.

Caffeine

Chocolate flavor contains 2.1 mg caffeine/8 fl oz.

E. ENSURE PLUS®

Usage: ENSURE PLUS is a high-calorie, low-residue liquid food for use when extra calories and nutrients, but a normal concentration of protein, are needed. It is designed primarily as an oral nutritional supplement to be used with or between meals or, in appropriate amounts, as a meal replacement. ENSURE PLUS is lactose- and gluten-free. Although it is primarily an oral nutritional supplement, it can be fed by tube.

Patient Conditions:

- For patients who require extra calories and nutrients, but a normal concentration of protein, in a limited volume
- For patients who need to gain or maintain healthy weight

Features

- Rich, creamy taste
- Good source of essential vitamins and minerals

Ingredients

Vanilla: ®-D Water, Corn Syrup, Maltodextrin (Corn), Corn Oil, Sodium and Calcium Caseinates, Sugar (Sucrose), Soy Protein Isolate, Magnesium Chloride,

Potassium Citrate, Calcium Phosphate Tribasic, Soy Lecithin, Natural and Artificial Flavor, Sodium Citrate, Potassium Chloride, Choline Chloride, Ascorbic Acid, Carrageenan, Zinc Sulfate, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Niacinamide, Calcium Pantothenate, Manganese Sulfate, Cupric Sulfate, Thiamine Chloride Hydrochloride, Pyridoxine Hydrochloride, Riboflavin, Vitamin A Palmitate, Folic Acid, Biotin, Chromium Chloride, Sodium Molybdate, Potassium Iodide, Sodium Selenite, Phylloquinone, Cyanocobalamin and Vitamin D_3.

Protein

The protein source is a blend of two high-biologic-value proteins: casein and soy.

Sodium and calcium caseinates	84%
Soy protein isolate	16%

Fat

The fat source is corn oil.

Corn oil	100%

Carbohydrate

ENSURE PLUS contains a combination of maltodextrin and sucrose. The mild sweetness and flavor variety (vanilla, chocolate, strawberry. coffee, buffer pecan, and eggnog), plus VARI-FLAVORS® Flavor Pacs in pecan, cherry, strawberry. lemon, and orange, help to prevent flavor fatigue and aid in patient compliance.

Vanilla, strawberry, butter pecan, and coffee flavors

Corn Syrup	39%
Maltodextrin	38%
Sucrose	23%

Chocolate and eggnog flavors

Corn Syrup	36%

Maltodextrin	34%
Sucrose	30%

Vitamins and Minerals

An 8-fl-oz serving of ENSURE PLUS provides at least 15% of the RDIs for 25 key Vitamins and minerals.

Caffeine

Chocolate flavor contains 3.1 mg Caffeine/8 fl oz. Coffee flavor contains a trace amount of caffeine.

F. ENSURE PLUS® HN

Usage: ENSURE PLUS HN is a nutritionally complete high-calorie, high-nitrogen liquid food designed for people with higher calorie and protein needs or limited volume tolerance. It may be used for oral supplementation or for total nutritional support by tube. ENSURE PLUS HN is lactose- and gluten-free.

Patient Conditions:

- For patients with increased calorie and protein needs, such as following surgery or injury
- For patients with limited volume tolerance and early satiety

Features

- For supplemental or total nutrition
- For oral or tube feeding
- 1.5 CaVmL
- High nitrogen
- Calorically dense

Ingredients

Vanilla: ®-D Water, Maltodextrin (Corn), Sodium and Calcium Caseinates, Corn Oil, Sugar (Sucrose), Soy Protein Isolate, Magnesium Chloride, Potassium Citrate, Calcium Phosphate Tribasic, Soy Lecithin, Natural and Artificial Flavor, Sodium Citrate, Choline Chloride, Ascorbic Acid, Taurine, L-Carnitine, Zinc Sulfate, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Niacinamide, Carrageenan, Calcium Pantothenate, Manganese Sulfate, Cupric Sulfate, Thiamine Chloride Hydrochloride, Pyridoxine Hydrochloride, Riboflavin, Vitamin A Palmitate, Folic Acid, Biotin, Chromium Chloride, Sodium Molybdate, Potassium Iodide, Sodium Selenite, Phylloquinone, Cyanocobalamin and Vitamin D_3.

G. ENSURE® POWDER

Usage: ENSURE POWDER (reconstituted with water) is a low-residue liquid food designed primarily as an oral nutritional supplement to be used with or between meals. ENSURE POWDER is lactose- and gluten-free, and is suitable for use in modified diets, including low-cholesterol diets.

Patient Conditions:

- For patients on modified diets
- For elderly patients at nutrition risk
- For patients recovering from illness/surgery
- For patients who need a low-residue diet

Features

- Convenient, easy to mix
- Low in saturated fat
- Contains 9 g of total fat and < 5 mg of cholesterol per serving
- High in vitamins and minerals
- For low-cholesterol diets
- Lactose-free, easily digested

Ingredients: ®-D Corn Syrup, Maltodextrin (Corn), Sugar (Sucrose), Corn Oil, Sodium and Calcium Caseinates, Soy Protein Isolate, Artificial Flavor, Potassium Citrate, Magnesium Chloride, Sodium Citrate, Calcium Phosphate Tribasic, Potassium Chloride, Soy Lecithin, Ascorbic Acid, Choline Chloride, Zinc Sulfate, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Niacinamide, Calcium Pantothenate, Manganese Sulfate, Thiamine Chloride Hydrochloride, Cupric Sulfate, Pyridoxine Hydrochloride, Riboflavin, Vitamin A Palmitate, Folic Acid, Biotin, Sodium Molybdate, Chromium Chloride, Potassium Iodide, Sodium Selenate, Phylloquinone, Vitamin D_3 and Cyanocobalamin.

Protein

The protein source is a blend of two high-biologic-value proteins: casein and soy.

Sodium and calcium caseinates	84%
Soy protein isolate	16%

Fat

The fat source is corn oil.

Corn oil	100%

Carbohydrate

ENSURE POWDER contains a combination of corn syrup, maltodextrin, and sucrose. The mild sweetness of ENSURE POWDER, plus VARI-FLAVORS® Flavor Pacs in pecan, cherry, strawberry, lemon, and orange, helps to prevent flavor fatigue and aid in patient compliance.

Vanilla

Corn Syrup	35%
Maltodextrin	35%
Sucrose	30%

H. ENSURE® PUDDING

Usage: ENSURE PUDDING is a nutrient-dense supplement providing balanced nutrition in a nonliquid form to be used with or between meals. It is appropriate for consistency-modified diets (e.g., soft, pureed, or full liquid) or for people with swallowing impairments. ENSURE PUDDING is gluten-free.

Patient Conditions:

- For patients on consistency-modified diets (e.g., soft, pureed, or full liquid)
- For patients with swallowing impairments

Features

- Rich and creamy, good taste
- Good source of essential vitamins and minerals Convenient-needs no refrigeration
- Gluten-free

Nutrient Profile per 5 oz: Calories 250, Protein 10.9%, Total Fat 34.9%, Carbohydrate 54.2%

Ingredients:

Vanilla: ®-D Nonfat Milk, Water, Sugar (Sucrose), Partially Hydrogenated Soybean Oil, Modified Food Starch, Magnesium Sulfate. Sodium Stearoyl Lactylate, Sodium Phosphate Dibasic, Artificial Flavor, Ascorbic Acid, Zinc Sulfate, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Choline Chloride, Niacinamide, Manganese Sulfate, Calcium Pantothenate, FD&C Yellow #5, Potassium Citrate, Cupric Sulfate, Vitamin A Palmitate, Thiamine Chloride Hydrochloride, Pyridoxine Hydrochloride, Riboflavin, FD&C Yellow #6, Folic Acid, Biotin, Phylloquinone, Vitamin D3 and Cyanocobalamin.

Protein

The protein source is nonfat milk.

Nonfat milk	100%

Fat

The fat source is hydrogenated soybean oil.

Hydrogenated soybean oil	100%

Carbohydrate

ENSURE PUDDING contains a combination of sucrose and modified food starch. The mild sweetness and flavor variety (vanilla, chocolate, butterscotch, and tapioca) help prevent flavor fatigue. The product contains 9.2 grams of lactose per serving.

Vanilla and other nonchocolate flavors

Sucrose	56%
Lactose	27%
Modified food starch	17%

Chocolate

Sucrose	58%
Lactose	26%
Modified food starch	16%

I. ENSURE® WITH FIBER

Usage: ENSURE WITH FIBER is a fiber-containing, nutritionally complete liquid food designed for people who can benefit from increased dietary fiber and nutrients. ENSURE WITH FIBER is suitable for people who do not require a low-residue diet. It can be fed orally or by tube, and can be used as a nutritional supplement to a regular diet or, in appropriate amounts, as a meal replacement. ENSURE WITH FIBER is lactose- and gluten-free, and is suitable for use in modified diets, including low-cholesterol diets.

Patient Conditions

- For patients who can benefit from increased dietary fiber and nutrients

Features

- New advanced formula-low in saturated fat, higher in vitamins and minerals
- Contains 6 g of total fat and < 5 mg of cholesterol per serving
- Rich, creamy taste
- Good source of fiber
- Excellent source of essential vitamins and minerals
- For low-cholesterol diets
- Lactose- and gluten-free

Ingredients

Vanilla: ®-D Water, Maltodextrin (Corn), Sugar (Sucrose), Sodium and Calcium Caseinates, Oat Fiber, High-Oleic Safflower Oil, Canola Oil, Soy Protein Isolate, Corn Oil, Soy Fiber, Calcium Phosphate Tribasic, Magnesium Chloride, Potassium Citrate, Cellulose Gel, Soy Lecithin, Potassium Phosphate Dibasic, Sodium Citrate, Natural and Artificial Flavors, Choline Chloride, Magnesium Phosphate, Ascorbic Acid, Cellulose Gum, Potassium Chloride, Carrageenan, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Zinc Sulfate, Niacinamide, Manganese Sulfate, Calcium Pantothenate, Cupric Sulfate, Vitamin A Palmitate, Thiamine Chloride Hydrochloride, Pyridoxine Hydrochloride, Riboflavin, Folic Acid, Chromium Chloride, Biotin, Sodium Molybdate, Potassium Iodide, Sodium Selenate, Phylloquinone, Vitamin D_3 and Cyanocobalamin.

Protein

The protein source is a blend of two high-biologic-value proteins- casein and soy.

Sodium and calcium caseinates	80%
Soy protein isolate	20%

Fat

The fat source is a blend of three oils: high-oleic safflower, canola, and corn.

High-oleic safflower oil	40%
Canola oil	40%
Corn oil	20%

The level of fat in ENSURE WITH FIBER meets American Heart Association (AHA) guidelines. The 6 grams of fat in ENSURE WITH FIBER represent 22% of the total calories, with 2.01 % of the fat being from saturated fatty acids and 6.7% from polyunsaturated fatty acids. These values are within the AHA guidelines of ≤ 30% of total calories from fat, < 1 0% of the calories from saturated fatty acids, and ≤ 1 0% of total calories from polyunsaturated fatty acids.

Carbohydrate

ENSURE WITH FIBER contains a combination of maltodextrin and sucrose. The mild sweetness and flavor variety (vanilla, chocolate, and butter pecan), plus VARI-FLAVORS® Flavor Pacs in pecan, cherry, strawberry, lemon, and orange, help to prevent flavor fatigue and aid in patient compliance.

Vanilla and other nonchocolate flavors

Maltodextrin	66%
Sucrose	25%
Oat Fiber	7%
Soy Fiber	2%

Chocolate

Maltodextrin	55%
Sucrose	36%
Oat Fiber	7%

Soy Fiber 2%

Fiber

The fiber blend used in ENSURE WITH FIBER consists of oat fiber and soy polysaccharide. This blend results in approximately 4 grams of total dietary fiber per 8-fl-oz can. The ratio of insoluble to soluble fiber is 95:5.

The various nutritional supplements described above and known to others of skill in the art can be substituted and/or supplemented with the PUFAs of this invention.

J. Oxepa™ Nutritional Product

Oxepa is low-carbohydrate, calorically dense enteral nutritional product designed for the dietary management of patients with or at risk for ARDS. It has a unique combination of ingredients, including a patented oil blend containing eicosapentaenoic acid (EPA from fish oil), γ-linolenic acid (GLA from borage oil), and elevated antioxidant levels.

Caloric Distribution:

- Caloric density is high at 1.5 Cal/mL (355 Cal/8 fl oz), to minimize the volume required to meet energy needs.
- The distribution of Calories in Oxepa is shown in Table 7.

Table 7. Caloric Distribution of Oxepa

	per 8 fl oz.	per liter	% of Cal
Calories	355	1,500	---
Fat (g)	22.2	93.7	55.2
Carbohydrate (g)	25	105.5	28.1
Protein (g)	14.8	62.5	16.7
Water (g)	186	785	---

Fat:

- Oxepa contains 22.2 g of fat per 8-fl oz serving (93.7 g/L).
- The fat source is a oil blend of 31.8% canola oil, 25% medium-chain triglycerides (MCTs), 20% borage oil, 20% fish oil, and 3.2 % soy lecithin. The typical fatty acid profile of Oxepa is shown in Table 8.

- Oxepa provides a balanced amount of polyunsaturated, monounsaturated, and saturated fatty acids, as shown in Table 10.
- Medium-chain trigylcerides (MCTs) -- 25% of the fat blend -- aid gastric emptying because they are absorbed by the intestinal tract without emulsification by bile acids.

The various fatty acid components of Oxepa™ nutritional product can be substituted and/or supplemented with the PUFAs of this invention.

Table 8. Typical Fatty Acid Profile

	% Total Fatty Acids	g/8 fl oz*	g/L*
Caproic (6:0)	0.2	0.04	0.18
Caprylic (8:0)	14.69	3.1	13.07
Capric (10:0)	11.06	2.33	9.87
Palmitic (16:0)	5.59	1.18	4.98
Palmitoleic (16:1n-7)	1.82	0.38	1.62
Stearic (18:0)	1.84	0.39	1.64
Oleic (18:1n-9)	24.44	5.16	21.75
Linoleic (18:2n-6)	16.28	3.44	14.49
α-Linolenic (18:3n-3)	3.47	0.73	3.09
γ-Linolenic (18:3n-6)	4.82	1.02	4.29
Eicosapentaenoic (20:5n-3)	5.11	1.08	4.55
n-3-Docosapentaenoic (22:5n-3)	0.55	0.12	0.49
Docosahexaenoic (22:6n-3)	2.27	0.48	2.02
Others	7.55	1.52	6.72

* Fatty acids equal approximately 95% of total fat.

Table 9. Fat Profile of Oxepa.

% of total calories from fat	55.2
Polyunsaturated fatty acids	31.44 g/L
Monounsaturated fatty acids	25.53 g/L
Saturated fatty acids	32.38 g/L
n-6 to n-3 ratio	1.75:1
Cholesterol	9.49 mg/8 fl oz 40.1 mg/L

Carbohydrate:

- The carbohydrate content is 25.0 g per 8-fl-oz serving (105.5 g/L).

- The carbohydrate sources are 45% maltodextrin (a complex carbohydrate) and 55% sucrose (a simple sugar), both of which are readily digested and absorbed.

- The high-fat and low-carbohydrate content of Oxepa is designed to minimize carbon dioxide (CO_2) production. High CO_2 levels can complicate weaning in ventilator-dependent patients. The low level of carbohydrate also may be useful for those patients who have developed stress-induced hyperglycemia.

- Oxepa is lactose-free.

Dietary carbohydrate, the amino acids from protein, and the glycerol moiety of fats can be converted to glucose within the body. Throughout this process, the carbohydrate requirements of glucose-dependent tissues (such as the central nervous system and red blood cells) are met. However, a diet free of carbohydrates can lead to ketosis, excessive catabolism of tissue protein, and loss of fluid and electrolytes. These effects can be prevented by daily ingestion of 50 to 100 g of digestible carbohydrate, if caloric intake is adequate. The carbohydrate level in Oxepa is also sufficient to minimize gluconeogenesis, if energy needs are being met.

Protein:

- Oxepa contains 14.8 g of protein per 8-fl-oz serving (62.5 g/L).

- The total calorie/nitrogen ratio (150:1) meets the need of stressed patients.

- Oxepa provides enough protein to promote anabolism and the maintenance of lean body mass without precipitating respiratory problems. High protein intakes are a concern in patients with respiratory insufficiency. Although protein has little effect on CO_2 production, a high protein diet will increase ventilatory drive.

- The protein sources of Oxepa are 86.8% sodium caseinate and 13.2% calcium caseinate.
- As demonstrated in Table 11, the amino acid profile of the protein system in Oxepa meets or surpasses the standard for high quality protein set by theNational Academy of Sciences.
- Oxepa is gluten-free.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: KNUTZON, DEBORAH
MURKERJI, PRADIP
HUANG, YUNG-SHENG
THURMOND, JENNIFER
CHAUDHARY, SUNITA
LEONARD, AMANDA

(ii) TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SYNTHESIS OF LONG CHAIN POLY-UNSATURATED FATTY ACIDS

(iii) NUMBER OF SEQUENCES: 40

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: LIMBACH AND LIMBACH LLP
(B) STREET: 2001 FERRY BUILDING
(C) CITY: SAN FRANCISCO
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 94111

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: Microsoft Word

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) (B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: WARD, MICHAEL R.
(B) REGISTRATION NUMBER: 38,651
(C) REFERENCE/DOCKET NUMBER: CGAB-210

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (415) 433-4150
(B) TELEFAX: (415) 433-8716
(C) TELEX: N/A

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1617 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGACACTCCT TCCTTCTTCT CACCCGTCCT AGTCCCCTTC AACCCCCCTC TTTGACAAAG 60

```
ACAACAAACC ATGGCTGCTG CTCCCAGTGT GAGGACGTTT ACTCGGGCCG AGGTTTTGAA    120
TGCCGAGGCT CTGAATGAGG GCAAGAAGGA TGCCGAGGCA CCCTTCTTGA TGATCATCGA    180
CAACAAGGTG TACGATGTCC GCGAGTTCGT CCCTGATCAT CCCGGTGGAA GTGTGATTCT    240
CACGCACGTT GGCAAGGACG GCACTGACGT CTTTGACACT TTTCACCCCG AGGCTGCTTG    300
GGAGACTCTT GCCAACTTTT ACGTTGGTGA TATTGACGAG AGCGACCGCG ATATCAAGAA    360
TGATGACTTT GCGGCCGAGG TCCGCAAGCT GCGTACCTTG TTCCAGTCTC TTGGTTACTA    420
CGATTCTTCC AAGGCATACT ACGCCTTCAA GGTCTCGTTC AACCTCTGCA TCTGGGGTTT    480
GTCGACGGTC ATTGTGGCCA AGTGGGGCCA GACCTCGACC CTCGCCAACG TGCTCTCGGC    540
TGCGCTTTTG GGTCTGTTCT GGCAGCAGTG CGGATGGTTG GCTCACGACT TTTTGCATCA    600
CCAGGTCTTC CAGGACCGTT TCTGGGGTGA TCTTTTCGGC GCCTTCTTGG GAGGTGTCTG    660
CCAGGGCTTC TCGTCCTCGT GGTGGAAGGA CAAGCACAAC ACTCACCACG CCGCCCCCAA    720
CGTCCACGGC GAGGATCCCG ACATTGACAC CCACCCTCTG TTGACCTGGA GTGAGCATGC    780
GTTGGAGATG TTCTCGGATG TCCCAGATGA GGAGCTGACC CGCATGTGGT CGCGTTTCAT    840
GGTCCTGAAC CAGACCTGGT TTTACTTCCC CATTCTCTCG TTTGCCCGTC TCTCCTGGTG    900
CCTCCAGTCC ATTCTCTTTG TGCTGCCTAA CGGTCAGGCC CACAAGCCCT CGGGCGCGCG    960
TGTGCCCATC TCGTTGGTCG AGCAGCTGTC GCTTGCGATG CACTGGACCT GGTACCTCGC   1020
CACCATGTTC CTGTTCATCA AGGATCCCGT CAACATGCTG GTGTACTTTT TGGTGTCGCA   1080
GGCGGTGTGC GGAAACTTGT TGGCGATCGT GTTCTCGCTC AACCACAACG GTATGCCTGT   1140
GATCTCGAAG GAGGAGGCGG TCGATATGGA TTTCTTCACG AAGCAGATCA TCACGGGTCG   1200
TGATGTCCAC CCGGGTCTAT TTGCCAACTG GTTCACGGGT GGATTGAACT ATCAGATCGA   1260
GCACCACTTG TTCCCTTCGA TGCCTCGCCA CAACTTTTCA AAGATCCAGC CTGCTGTCGA   1320
GACCCTGTGC AAAAAGTACA ATGTCCGATA CCACACCACC GGTATGATCG AGGGAACTGC   1380
AGAGGTCTTT AGCCGTCTGA ACGAGGTCTC CAAGGCTGCC TCCAAGATGG GTAAGGCGCA   1440
GTAAAAAAAA AAACAAGGAC GTTTTTTTTC GCCAGTGCCT GTGCCTGTGC CTGCTTCCCT   1500
TGTCAAGTCG AGCGTTTCTG GAAAGGATCG TTCAGTGCAG TATCATCATT CTCCTTTTAC   1560
CCCCCGCTCA TATCTCATTC ATTTCTCTTA TTAAACAACT TGTTCCCCCC TTCACCG      1617
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 457 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ala Ala Ala Pro Ser Val Arg Thr Phe Thr Arg Ala Glu Val Leu
1               5                   10                  15

Asn Ala Glu Ala Leu Asn Glu Gly Lys Lys Asp Ala Glu Ala Pro Phe
            20                  25                  30

Leu Met Ile Ile Asp Asn Lys Val Tyr Asp Val Arg Glu Phe Val Pro
        35                  40                  45

Asp His Pro Gly Gly Ser Val Ile Leu Thr His Val Gly Lys Asp Gly
    50                  55                  60

Thr Asp Val Phe Asp Thr Phe His Pro Glu Ala Ala Trp Glu Thr Leu
65                  70                  75                  80

Ala Asn Phe Tyr Val Gly Asp Ile Asp Glu Ser Asp Arg Asp Ile Lys
                85                  90                  95

Asn Asp Asp Phe Ala Ala Glu Val Arg Lys Leu Arg Thr Leu Phe Gln
            100                 105                 110

Ser Leu Gly Tyr Tyr Asp Ser Ser Lys Ala Tyr Tyr Ala Phe Lys Val
        115                 120                 125

Ser Phe Asn Leu Cys Ile Trp Gly Leu Ser Thr Val Ile Val Ala Lys
    130                 135                 140

Trp Gly Gln Thr Ser Thr Leu Ala Asn Val Leu Ser Ala Ala Leu Leu
145                 150                 155                 160

Gly Leu Phe Trp Gln Gln Cys Gly Trp Leu Ala His Asp Phe Leu His
                165                 170                 175

His Gln Val Phe Gln Asp Arg Phe Trp Gly Asp Leu Phe Gly Ala Phe
            180                 185                 190

Leu Gly Gly Val Cys Gln Gly Phe Ser Ser Ser Trp Trp Lys Asp Lys
        195                 200                 205

His Asn Thr His His Ala Ala Pro Asn Val His Gly Glu Asp Pro Asp
    210                 215                 220

Ile Asp Thr His Pro Leu Leu Thr Trp Ser Glu His Ala Leu Glu Met
225                 230                 235                 240

Phe Ser Asp Val Pro Asp Glu Glu Leu Thr Arg Met Trp Ser Arg Phe
                245                 250                 255

Met Val Leu Asn Gln Thr Trp Phe Tyr Phe Pro Ile Leu Ser Phe Ala
            260                 265                 270

Arg Leu Ser Trp Cys Leu Gln Ser Ile Leu Phe Val Leu Pro Asn Gly
        275                 280                 285

Gln Ala His Lys Pro Ser Gly Ala Arg Val Pro Ile Ser Leu Val Glu
    290                 295                 300

Gln Leu Ser Leu Ala Met His Trp Thr Trp Tyr Leu Ala Thr Met Phe
305                 310                 315                 320

Leu Phe Ile Lys Asp Pro Val Asn Met Leu Val Tyr Phe Leu Val Ser
                325                 330                 335

Gln Ala Val Cys Gly Asn Leu Leu Ala Ile Val Phe Ser Leu Asn His
```

```
                340                                345                                350

Asn Gly Met Pro Val Ile Ser Lys Glu Glu Ala Val Asp Met Asp Phe
        355                 360                 365

Phe Thr Lys Gln Ile Ile Thr Gly Arg Asp Val His Pro Gly Leu Phe
    370                 375                 380

Ala Asn Trp Phe Thr Gly Gly Leu Asn Tyr Gln Ile Glu His His Leu
385                 390                 395                 400

Phe Pro Ser Met Pro Arg His Asn Phe Ser Lys Ile Gln Pro Ala Val
                405                 410                 415

Glu Thr Leu Cys Lys Lys Tyr Asn Val Arg Tyr His Thr Thr Gly Met
            420                 425                 430

Ile Glu Gly Thr Ala Glu Val Phe Ser Arg Leu Asn Glu Val Ser Lys
        435                 440                 445

Ala Ala Ser Lys Met Gly Lys Ala Gln
    450                 455
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1488 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
GTCCCCTGTC GCTGTCGGCA CACCCCATCC TCCCTCGCTC CCTCTGCGTT TGTCCTTGGC      60
CCACCGTCTC TCCTCCACCC TCCGAGACGA CTGCAACTGT AATCAGGAAC CGACAAATAC     120
ACGATTTCTT TTTACTCAGC ACCAACTCAA AATCCTCAAC CGCAACCCTT TTTCAGGATG     180
GCACCTCCCA ACACTATCGA TGCCGGTTTG ACCCAGCGTC ATATCAGCAC CTCGGCCCCA     240
AACTCGGCCA AGCCTGCCTT CGAGCGCAAC TACCAGCTCC CCGAGTTCAC CATCAAGGAG     300
ATCCGAGAGT GCATCCCTGC CCACTGCTTT GAGCGCTCCG GTCTCCGTGG TCTCTGCCAC     360
GTTGCCATCG ATCTGACTTG GGCGTCGCTC TTGTTCCTGG CTGCGACCCA GATCGACAAG     420
TTTGAGAATC CCTTGATCCG CTATTTGGCC TGGCCTGTTT ACTGGATCAT GCAGGGTATT     480
GTCTGCACCG GTGTCTGGGT GCTGGCTCAC GAGTGTGGTC ATCAGTCCTT CTCGACCTCC     540
AAGACCCTCA ACAACACAGT TGGTTGGATC TTGCACTCGA TGCTCTTGGT CCCCTACCAC     600
TCCTGGAGAA TCTCGCACTC GAAGCACCAC AAGGCCACTG GCCATATGAC CAAGGACCAG     660
GTCTTTGTGC CCAAGACCCG CTCCCAGGTT GGCTTGCCTC CCAAGGAGAA CGCTGCTGCT     720
GCCGTTCAGG AGGAGGACAT GTCCGTGCAC CTGGATGAGG AGGCTCCCAT TGTGACTTTG     780
TTCTGGATGG TGATCCAGTT CTTGTTCGGA TGGCCCGCGT ACCTGATTAT GAACGCCTCT     840
```

```
GGCCAAGACT ACGGCCGCTG GACCTCGCAC TTCCACACGT ACTCGCCCAT CTTTGAGCCC     900
CGCAACTTTT TCGACATTAT TATCTCGGAC CTCGGTGTGT TGGCTGCCCT CGGTGCCCTG     960
ATCTATGCCT CCATGCAGTT GTCGCTCTTG ACCGTCACCA AGTACTATAT TGTCCCCTAC    1020
CTCTTTGTCA ACTTTTGGTT GGTCCTGATC ACCTTCTTGC AGCACACCGA TCCCAAGCTG    1080
CCCCATTACC GCGAGGGTGC CTGGAATTTC CAGCGTGGAG CTCTTTGCAC CGTTGACCGC    1140
TCGTTTGGCA AGTTCTTGGA CCATATGTTC CACGGCATTG TCCACACCCA TGTGGCCCAT    1200
CACTTGTTCT CGCAAATGCC GTTCTACCAT GCTGAGGAAG CTACCTATCA TCTCAAGAAA    1260
CTGCTGGGAG AGTACTATGT GTACGACCCA TCCCCGATCG TCGTTGCGGT CTGGAGGTCG    1320
TTCCGTGAGT GCCGATTCGT GGAGGATCAG GGAGACGTGG TCTTTTTCAA GAAGTAAAAA    1380
AAAAGACAAT GGACCACACA CAACCTTGTC TCTACAGACC TACGTATCAT GTAGCCATAC    1440
CACTTCATAA AAGAACATGA GCTCTAGAGG CGTGTCATTC GCGCCTCC                 1488
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 399 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Ala Pro Pro Asn Thr Ile Asp Ala Gly Leu Thr Gln Arg His Ile
1               5                   10                  15

Ser Thr Ser Ala Pro Asn Ser Ala Lys Pro Ala Phe Glu Arg Asn Tyr
            20                  25                  30

Gln Leu Pro Glu Phe Thr Ile Lys Glu Ile Arg Glu Cys Ile Pro Ala
        35                  40                  45

His Cys Phe Glu Arg Ser Gly Leu Arg Gly Leu Cys His Val Ala Ile
    50                  55                  60

Asp Leu Thr Trp Ala Ser Leu Leu Phe Leu Ala Ala Thr Gln Ile Asp
65                  70                  75                  80

Lys Phe Glu Asn Pro Leu Ile Arg Tyr Leu Ala Trp Pro Val Tyr Trp
                85                  90                  95

Ile Met Gln Gly Ile Val Cys Thr Gly Val Trp Val Leu Ala His Glu
            100                 105                 110

Cys Gly His Gln Ser Phe Ser Thr Ser Lys Thr Leu Asn Asn Thr Val
        115                 120                 125

Gly Trp Ile Leu His Ser Met Leu Leu Val Pro Tyr His Ser Trp Arg
    130                 135                 140

Ile Ser His Ser Lys His His Lys Ala Thr Gly His Met Thr Lys Asp
145                 150                 155                 160
```

```
Gln Val Phe Val Pro Lys Thr Arg Ser Gln Val Gly Leu Pro Pro Lys
                165                 170                 175

Glu Asn Ala Ala Ala Ala Val Gln Glu Glu Asp Met Ser Val His Leu
            180                 185                 190

Asp Glu Glu Ala Pro Ile Val Thr Leu Phe Trp Met Val Ile Gln Phe
        195                 200                 205

Leu Phe Gly Trp Pro Ala Tyr Leu Ile Met Asn Ala Ser Gly Gln Asp
    210                 215                 220

Tyr Gly Arg Trp Thr Ser His Phe His Thr Tyr Ser Pro Ile Phe Glu
225                 230                 235                 240

Pro Arg Asn Phe Phe Asp Ile Ile Ile Ser Asp Leu Gly Val Leu Ala
                245                 250                 255

Ala Leu Gly Ala Leu Ile Tyr Ala Ser Met Gln Leu Ser Leu Leu Thr
            260                 265                 270

Val Thr Lys Tyr Tyr Ile Val Pro Tyr Leu Phe Val Asn Phe Trp Leu
        275                 280                 285

Val Leu Ile Thr Phe Leu Gln His Thr Asp Pro Lys Leu Pro His Tyr
    290                 295                 300

Arg Glu Gly Ala Trp Asn Phe Gln Arg Gly Ala Leu Cys Thr Val Asp
305                 310                 315                 320

Arg Ser Phe Gly Lys Phe Leu Asp His Met Phe His Gly Ile Val His
                325                 330                 335

Thr His Val Ala His His Leu Phe Ser Gln Met Pro Phe Tyr His Ala
            340                 345                 350

Glu Glu Ala Thr Tyr His Leu Lys Lys Leu Leu Gly Glu Tyr Tyr Val
        355                 360                 365

Tyr Asp Pro Ser Pro Ile Val Val Ala Val Trp Arg Ser Phe Arg Glu
    370                 375                 380

Cys Arg Phe Val Glu Asp Gln Gly Asp Val Val Phe Phe Lys Lys
385                 390                 395
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 355 amino acids
  (B) TYPE: amino acid
  (C) STRANDEDNESS: not relevant
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Glu Val Arg Lys Leu Arg Thr Leu Phe Gln Ser Leu Gly Tyr Tyr Asp
1               5                   10                  15

Ser Ser Lys Ala Tyr Tyr Ala Phe Lys Val Ser Phe Asn Leu Cys Ile
            20                  25                  30
```

```
Trp Gly Leu Ser Thr Val Ile Val Ala Lys Trp Gly Gln Thr Ser Thr
        35                  40                  45

Leu Ala Asn Val Leu Ser Ala Ala Leu Leu Gly Leu Phe Trp Gln Gln
    50                  55                  60

Cys Gly Trp Leu Ala His Asp Phe Leu His His Gln Val Phe Gln Asp
65                  70                  75                  80

Arg Phe Trp Gly Asp Leu Phe Gly Ala Phe Leu Gly Gly Val Cys Gln
                85                  90                  95

Gly Phe Ser Ser Ser Trp Trp Lys Asp Lys His Asn Thr His His Ala
            100                 105                 110

Ala Pro Asn Val His Gly Glu Asp Pro Asp Ile Asp Thr His Pro Leu
        115                 120                 125

Leu Thr Trp Ser Glu His Ala Leu Glu Met Phe Ser Asp Val Pro Asp
    130                 135                 140

Glu Glu Leu Thr Arg Met Trp Ser Arg Phe Met Val Leu Asn Gln Thr
145                 150                 155                 160

Trp Phe Tyr Phe Pro Ile Leu Ser Phe Ala Arg Leu Ser Trp Cys Leu
                165                 170                 175

Gln Ser Ile Leu Phe Val Leu Pro Asn Gly Gln Ala His Lys Pro Ser
            180                 185                 190

Gly Ala Arg Val Pro Ile Ser Leu Val Glu Gln Leu Ser Leu Ala Met
        195                 200                 205

His Trp Thr Trp Tyr Leu Ala Thr Met Phe Leu Phe Ile Lys Asp Pro
    210                 215                 220

Val Asn Met Leu Val Tyr Phe Leu Val Ser Gln Ala Val Cys Gly Asn
225                 230                 235                 240

Leu Leu Ala Ile Val Phe Ser Leu Asn His Asn Gly Met Pro Val Ile
                245                 250                 255

Ser Lys Glu Glu Ala Val Asp Met Asp Phe Phe Thr Lys Gln Ile Ile
            260                 265                 270

Thr Gly Arg Asp Val His Pro Gly Leu Phe Ala Asn Trp Phe Thr Gly
        275                 280                 285

Gly Leu Asn Tyr Gln Ile Glu His His Leu Phe Pro Ser Met Pro Arg
    290                 295                 300

His Asn Phe Ser Lys Ile Gln Pro Ala Val Glu Thr Leu Cys Lys Lys
305                 310                 315                 320

Tyr Asn Val Arg Tyr His Thr Thr Gly Met Ile Glu Gly Thr Ala Glu
                325                 330                 335

Val Phe Ser Arg Leu Asn Glu Val Ser Lys Ala Ala Ser Lys Met Gly
            340                 345                 350

Lys Ala Gln
        355
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 104 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Val Thr Leu Tyr Thr Leu Ala Phe Val Ala Ala Asn Ser Leu Gly Val
1               5                   10                  15

Leu Tyr Gly Val Leu Ala Cys Pro Ser Val Xaa Pro His Gln Ile Ala
            20                  25                  30

Ala Gly Leu Leu Gly Leu Leu Trp Ile Gln Ser Ala Tyr Ile Gly Xaa
        35                  40                  45

Asp Ser Gly His Tyr Val Ile Met Ser Asn Lys Ser Asn Asn Xaa Phe
    50                  55                  60

Ala Gln Leu Leu Ser Gly Asn Cys Leu Thr Gly Ile Ile Ala Trp Trp
65                  70                  75                  80

Lys Trp Thr His Asn Ala His His Leu Ala Cys Asn Ser Leu Asp Tyr
                85                  90                  95

Gly Pro Asn Leu Gln His Ile Pro
            100
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 252 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Gly Val Leu Tyr Gly Val Leu Ala Cys Thr Ser Val Phe Ala His Gln
1               5                   10                  15

Ile Ala Ala Ala Leu Leu Gly Leu Leu Trp Ile Gln Ser Ala Tyr Ile
            20                  25                  30

Gly His Asp Ser Gly His Tyr Val Ile Met Ser Asn Lys Ser Tyr Asn
        35                  40                  45

Arg Phe Ala Gln Leu Leu Ser Gly Asn Cys Leu Thr Gly Ile Ser Ile
    50                  55                  60

Ala Trp Trp Lys Trp Thr His Asn Ala His His Leu Ala Cys Asn Ser
65                  70                  75                  80

Leu Asp Tyr Asp Pro Asp Leu Gln His Ile Pro Val Phe Ala Val Ser
                85                  90                  95
```

Thr Lys Phe Phe Ser Ser Leu Thr Ser Arg Phe Tyr Asp Arg Lys Leu
100 105 110

Thr Phe Gly Pro Val Ala Arg Phe Leu Val Ser Tyr Gln His Phe Thr
115 120 125

Tyr Tyr Pro Val Asn Cys Phe Gly Arg Ile Asn Leu Phe Ile Gln Thr
130 135 140

Phe Leu Leu Leu Phe Ser Lys Arg Glu Val Pro Asp Arg Ala Leu Asn
145 150 155 160

Phe Ala Gly Ile Leu Val Phe Trp Thr Trp Phe Pro Leu Leu Val Ser
165 170 175

Cys Leu Pro Asn Trp Pro Glu Arg Phe Phe Phe Val Phe Thr Ser Phe
180 185 190

Thr Val Thr Ala Leu Gln His Ile Gln Phe Thr Leu Asn His Phe Ala
195 200 205

Ala Asp Val Tyr Val Gly Pro Pro Thr Gly Ser Asp Trp Phe Glu Lys
210 215 220

Gln Ala Ala Gly Thr Ile Asp Ile Ser Cys Arg Ser Tyr Met Asp Trp
225 230 235 240

Phe Phe Gly Gly Leu Gln Phe Gln Leu Glu His His
245 250

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 125 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gly Xaa Xaa Asn Phe Ala Gly Ile Leu Val Phe Trp Thr Trp Phe Pro
1 5 10 15

Leu Leu Val Ser Cys Leu Pro Asn Trp Pro Glu Arg Phe Xaa Phe Val
20 25 30

Phe Thr Gly Phe Thr Val Thr Ala Leu Gln His Ile Gln Phe Thr Leu
35 40 45

Asn His Phe Ala Ala Asp Val Tyr Val Gly Pro Pro Thr Gly Ser Asp
50 55 60

Trp Phe Glu Lys Gln Ala Ala Gly Thr Ile Asp Ile Ser Cys Arg Ser
65 70 75 80

Tyr Met Asp Trp Phe Phe Cys Gly Leu Gln Phe Gln Leu Glu His His
85 90 95

Leu Phe Pro Arg Leu Pro Arg Cys His Leu Arg Lys Val Ser Pro Val
100 105 110

```
Gly Gln Arg Gly Phe Gln Arg Lys Xaa Asn Leu Ser Xaa
          115                 120                 125
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 131 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
Pro Ala Thr Glu Val Gly Gly Leu Ala Trp Met Ile Thr Phe Tyr Val
1               5                   10                  15

Arg Phe Phe Leu Thr Tyr Val Pro Leu Leu Gly Leu Lys Ala Phe Leu
            20                  25                  30

Gly Leu Phe Phe Ile Val Arg Phe Leu Glu Ser Asn Trp Phe Val Trp
        35                  40                  45

Val Thr Gln Met Asn His Ile Pro Met His Ile Asp His Asp Arg Asn
    50                  55                  60

Met Asp Trp Val Ser Thr Gln Leu Gln Ala Thr Cys Asn Val His Lys
65                  70                  75                  80

Ser Ala Phe Asn Asp Trp Phe Ser Gly His Leu Asn Phe Gln Ile Glu
                85                  90                  95

His His Leu Phe Pro Thr Met Pro Arg His Asn Tyr His Xaa Val Ala
            100                 105                 110

Pro Leu Val Gln Ser Leu Cys Ala Lys His Gly Ile Glu Tyr Gln Ser
        115                 120                 125

Lys Pro Leu
    130
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 87 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Cys Ser Pro Lys Ser Ser Pro Thr Arg Asn Met Thr Pro Ser Pro Phe
1               5                   10                  15

Ile Asp Trp Leu Trp Gly Gly Leu Asn Tyr Gln Ile Glu His His Leu
            20                  25                  30
```

```
Phe Pro Thr Met Pro Arg Cys Asn Leu Asn Arg Cys Met Lys Tyr Val
        35                  40                      45

Lys Glu Trp Cys Ala Glu Asn Asn Leu Pro Tyr Leu Val Asp Asp Tyr
    50                  55                  60

Phe Val Gly Tyr Asn Leu Asn Leu Gln Gln Leu Lys Asn Met Ala Glu
65                  70                  75                  80

Leu Val Gln Ala Lys Ala Ala
                85
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 143 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
Arg His Glu Ala Ala Arg Gly Gly Thr Arg Leu Ala Tyr Met Leu Val
1               5                   10                  15

Cys Met Gln Trp Thr Asp Leu Leu Trp Ala Ala Ser Phe Tyr Ser Arg
            20                  25                  30

Phe Phe Leu Ser Tyr Ser Pro Phe Tyr Gly Ala Thr Gly Thr Leu Leu
        35                  40                  45

Leu Phe Val Ala Val Arg Val Leu Glu Ser His Trp Phe Val Trp Ile
    50                  55                  60

Thr Gln Met Asn His Ile Pro Lys Glu Ile Gly His Glu Lys His Arg
65                  70                  75                  80

Asp Trp Ala Ser Ser Gln Leu Ala Ala Thr Cys Asn Val Glu Pro Ser
                85                  90                  95

Leu Phe Ile Asp Trp Phe Ser Gly His Leu Asn Phe Gln Ile Glu His
            100                 105                 110

His Leu Phe Pro Thr Met Thr Arg His Asn Tyr Arg Xaa Val Ala Pro
        115                 120                 125

Leu Val Lys Ala Phe Cys Ala Lys His Gly Leu His Tyr Glu Val
    130                 135                 140
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCAAGCTTCT GCAGGAGCTC TTTTTTTTTT TTTTT 35

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CUACUACUAC UAGGAGTCCT CTACGGTGTT TTG 33

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CAUCAUCAUC AUATGATGCT CAAGCTGAAA CTG 33

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TACCAACTCG AGAAAATGGC TGCTGCTCCC AGTGTGAGG 39

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AACTGATCTA GATTACTGCG CCTTACCCAT CTTGGAGGC 39

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TACCAACTCG AGAAAATGGC ACCTCCCAAC ACTATCGAT 39

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

AACTGATCTA GATTACTTCT TGAAAAAGAC CACGTCTCC 39

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 746 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CGTATGTCAC TCCATTCCAA ACTCGTTCAT GGTATCATAA ATATCAACAC ATTTACGCTC 60
CACTCCTCTA TGGTATTTAC ACACTCAAAT ATCGTACTCA AGATTGGGAA GCTTTTGTAA 120
AGGATGGTAA AAATGGTGCA ATTCGTGTTA GTGTCGCCAC AAATTTCGAT AAGGCCGCTT 180
ACGTCATTGG TAAATTGTCT TTTGTTTTCT TCCGTTTCAT CCTTCCACTC CGTTATCATA 240
GCTTTACAGA TTTAATTTGT TATTTCCTCA TTGCTGAATT CGTCTTTGGT TGGTATCTCA 300
CAATTAATTT CCAAGTTAGT CATGTCGCTG AAGATCTCAA ATTCTTTGCT ACCCCTGAAA 360
GACCAGATGA ACCATCTCAA ATCAATGAAG ATTGGGCAAT CCTTCAACTT AAAACTACTC 420
AAGATTATGG TCATGGTTCA CTCCTTTGTA CCTTTTTTAG TGGTTCTTTA AATCATCAAG 480
TTGTTCATCA TTTATTCCCA TCAATTGCTC AAGATTTCTA CCCACAACTT GTACCAATTG 540
TAAAAGAAGT TTGTAAAGAA CATAACATTA CTTACCACAT TAAACCAAAC TTCACTGAAG 600
CTATTATGTC ACACATTAAT TACCTTTACA AAATGGGTAA TGATCCAGAT TATGTTAAAA 660
AACCATTAGC CTCAAAAGAT GATTAAATGA AATAACTTAA AAACCAATTA TTTACTTTTG 720

ACAAACAGTA ATATTAATAA ATACAA 746

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 227 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
Tyr Val Thr Pro Phe Gln Thr Arg Ser Trp Tyr His Lys Tyr Gln
1               5                   10                  15
His Ile Tyr Ala Pro Leu Leu Tyr Gly Ile Tyr Thr Leu Lys Tyr
                20                  25                  30
Arg Thr Gln Asp Trp Glu Ala Phe Val Lys Asp Gly Lys Asn Gly
                35                  40                  45
Ala Ile Arg Val Ser Val Ala Thr Asn Phe Asp Lys Ala Ala Tyr
                50                  55                  60
Val Ile Gly Lys Leu Ser Phe Val Phe Phe Arg Phe Ile Leu Pro
                65                  70                  75
Leu Arg Tyr His Ser Phe Thr Asp Leu Ile Cys Tyr Phe Leu Ile
                80                  85                  90
Ala Glu Phe Val Phe Gly Trp Tyr Leu Thr Ile Asn Phe Gln Val
                95                  100                 105
Ser His Val Ala Glu Asp Leu Lys Phe Phe Ala Thr Pro Glu Arg
                110                 115                 120
Pro Asp Glu Pro Ser Gln Ile Asn Glu Asp Trp Ala Ile Leu Gln
                125                 130                 135
Leu Lys Thr Thr Gln Asp Tyr Gly His Gly Ser Leu Leu Cys Thr
                140                 145                 150
Phe Phe Ser Gly Ser Leu Asn His Gln Val Val His His Leu Phe
                155                 160                 165
Pro Ser Ile Ala Gln Asp Phe Tyr Pro Gln Leu Val Pro Ile Val
                170                 175                 180
Lys Glu Val Cys Lys Glu His Asn Ile Thr Tyr His Ile Lys Pro
                185                 190                 195
Asn Phe Thr Glu Ala Ile Met Ser His Ile Asn Tyr Leu Tyr Lys
                200                 205                 210
Met Gly Asn Asp Pro Asp Tyr Val Lys Lys Pro Leu Ala Ser Lys
                215                 220                 225
Asp Asp ***
```

(2) INFORMATION FOR SEQ ID NO 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 494 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid

(xi) SEQUENCE DESCRIPTION:SEQ ID NO:21:

```
TTTTGGAAGG NTCCAAGTTN ACCACGGANT NGGCAAGTTN ACGGGGCGGA AANCGGTTTT    60
CCCCCCAAGC CTTTTGTCGA CTGGTTCTGT GGTGGCTTCC AGTACCAAGT CGACCACCAC   120
TTATTCCCCA GCCTGCCCCG ACACAATCTG GCCAAGACAC ACGCACTGGT CGAATCGTTC   180
TGCAAGGAGT GGGGTGTCCA GTACCACGAA GCCGACCTCG TGGACGGGAC CATGGAAGTC   240
TTGCACCATT TGGGCAGCGT GGCCGGCGAA TTCGTCGTGG ATTTTGTACG CGACGGACCC   300
```

```
GCCATGTAAT CGTCGTTCGT GACGATGCAA GGGTTCACGC ACATCTACAC ACACTCACTC   360
ACACAACTAG TGTAACTCGT ATAGAATTCG GTGTCGACCT GGACCTTGTT TGACTGGTTG   420
GGGATAGGGT AGGTAGGCGG ACGCGTGGGT CGNCCCCGGG AATTCTGTGA CCGGTACCTG   480
GCCCGCGTNA AAGT                                                     494
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 87 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
Phe Trp Lys Xxx Pro Ser Xxx Pro Arg Xxx Xxx Gln Val Xxx Gly
1               5                   10                  15
Ala Glu Xxx Gly Phe Pro Pro Lys Pro Phe Val Asp Trp Phe Cys
                20                  25                  30
Gly Gly Phe Gln Tyr Gln Val Asp His His Leu Phe Pro Ser Leu
                35                  40                  45
Pro Arg His Asn Leu Ala Lys Thr His Ala Leu Val Glu Ser Phe
                50                  55                  60
Cys Lys Glu Trp Gly Val Gln Tyr His Glu Ala Asp Leu Val Asp
                65                  70                  75
Gly Thr Met Glu Val Leu His His Leu Gly Ser Val Ala Gly Glu
                65                  70                  75
Phe Val Val Asp Phe Val Arg Asp Gly Pro Ala Met
                80                  85
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 520 nucleic acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
GGATGGAGTT CGTCTGGATC GCTGTGCGCT ACGCGACGTG GTTTAAGCGT CATGGGTGCG    60
CTTGGGTACA CGCCGGGGCA GTCGTTGGGC ATGTACTTGT GCGCCTTTGG TCTCGGCTGC   120
ATTTACATTT TTCTGCAGTT CGCCGTAAGT CACACCCATT TGCCCGTGAG CAACCCGGAG   180
GATCAGCTGC ATTGGCTCGA GTACGCGCGG ACCACACTGT GAACATCAGC ACCAAGTCGT   240
GGTTTGTCAC ATGGTGGATG TCGAACCTCA ACTTTCAGAT CGAGCACCAC CTTTTCCCCA   300
CGGCGCCCCA GTTCCGTTTC AAGGAGATCA GCCCGCGCGT CGAGGCCCTC TTCAAGCGCC   360
ACGGTCTCCC TTACTACGAC ATGCCCTACA CGAGCGCCGT CTCCACCACC TTTGCCAACC   420
TCTACTCCGT CGGCCATTCC GTCGGCGACG CCAAGCGCGA CTAGCCTCTT TTCCTAGACC   480
TTAATTCCCC ACCCCACCCC ATGTTCTGTC TTCCTCCCGC                         520
```

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 153 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
Met Glu Phe Val Trp Ile Ala Val Arg Tyr Ala Thr Trp Phe Lys
1               5                   10                  15
Arg His Gly Cys Ala Trp Val His Ala Gly Ala Val Val Gly His
                20                  25                  30
Val Leu Val Arg Leu Trp Ser Arg Leu His Leu His Phe Ser Ala
                35                  40                  45
Val Arg Arg Lys Ser His Pro Phe Ala Arg Glu Gln Pro Gly Gly
                50                  55                  60
Ser Ala Ala Leu Ala Arg Val Arg Ala Asp His Thr Val Asn Ile
                65                  70                  75
Ser Thr Lys Ser Trp Phe Val Thr Trp Trp Met Ser Asn Leu Asn
                80                  85                  90
Phe Gln Ile Glu His His Leu Phe Pro Thr Ala Pro Gln Phe Arg
                95                  100                 105
Phe Lys Glu Ile Ser Pro Arg Val Glu Ala Leu Phe Lys Arg His
                110                 115                 120
Gly Leu Pro Tyr Tyr Asp Met Pro Tyr Thr Ser Ala Val Ser Thr
                125                 130                 135
Thr Phe Ala Asn Leu Tyr Ser Val Gly His Ser Val Gly Asp Ala
                140                 145                 150
Lys Arg Asp
```

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 420 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
ACGCGTCCGC CCACGCGTCC GCCGCGAGCA ACTCATCAAG GAAGGCTACT TTGACCCCTC    60
GCTCCCGCAC ATGACGTACC GCGTGGTCGA GATTGTTGTT CTCTTCGTGC TTTCCTTTTG   120
GCTGATGGGT CAGTCTTCAC CCCTCGCGCT CGCTCTCGGC ATTGTCGTCA GCGGCATCTC   180
TCAGGGTCGC TGCGGCTGGG TAATGCATGA GATGGGCCAT GGGTCGTTCA CTGGTGTCAT   240
TTGGCTTGAC GACCGGTTGT GCGAGTTCTT TTACGGCGTT GGTTGTGGCA TGAGCGGTCA   300
TTACTGGAAA AACCAGCACA GCAAACACCA CGCAGCGCCA AACCGGCTCG AGCACGATGT   360
AGATCTCAAC ACCTTGCCAT TGGTGGCCTT CAACGAGCGC GTCGTGCGCA AGGTCCGACC   420
```

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 125 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
Arg Val Arg Pro Arg Val Arg Arg Glu Gln Leu Ile Lys Glu Gly
1               5                   10                  15
Tyr Phe Asp Pro Ser Leu Pro His Met Thr Tyr Arg Val Val Glu
                20                  25                  30
Ile Val Val Leu Phe Val Leu Ser Phe Trp Leu Met Gly Gln Ser
                35                  40                  45
Ser Pro Leu Ala Leu Ala Leu Gly Ile Val Val Ser Gly Ile Ser
                50                  55                  60
Gln Gly Arg Cys Gly Trp Val Met His Glu Met Gly His Gly Ser
                65                  70                  75
Phe Thr Gly Val Ile Trp Leu Asp Asp Arg Leu Cys Glu Phe Phe
                65                  70                  75
Tyr Gly Val Gly Cys Gly Met Ser Gly His Tyr Trp Lys Asn Gln
                80                  85                  90
His Ser Lys His His Ala Ala Pro Asn Arg Leu Glu His Asp Val
                95                  100                 105
Asp Leu Asn Thr Leu Pro Leu Val Ala Phe Asn Glu Arg Val Val
                110                 115                 120
Arg Lys Val Arg Pro
                125
```

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1219 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Edited Contig 2692004)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
GCACGCCGAC CGGCGCCGGG AGATCCTGGC AAAGTATCCA GAGATAAAGT CCTTGATGAA    60
ACCTGATCCC AATTTGATAT GGATTATAAT TATGATGGTT CTCACCCAGT TGGGTGCATT   120
TTACATAGTA AAAGACTTGG ACTGGAAATG GGTCATATTT GGGGCCTATG CGTTTGGCAG   180
TTGCATTAAC CACTCAATGA CTCTGGCTAT TCATGAGATT GCCCACAATG CTGCCTTTGG   240
CAACTGCAAA GCAATGTGGA ATCGCTGGTT TGGAATGTTT GCTAATCTTC CTATTGGGAT   300
TCCATATTCA ATTTCCTTTA AGAGGTATCA CATGGATCAT CATCGGTACC TTGGAGCTGA   360
TGGCGTCGAT GTAGATATTC CTACCGATTT TGAGGGCTGG TTCTTCTGTA CCGCTTTCAG   420
AAAGTTTATA TGGGTTATTC TTCAGCCTCT CTTTTATGCC TTTCGACCTC TGTTCATCAA   480
CCCCAAACCA ATTACGTATC TGGAAGTTAT CAATACCGTG GCACAGGTCA CTTTTGACAT   540
TTTAATTTAT TACTTTTTGG GAATTAAATC CTTAGTCTAC ATGTTGGCAG CATCTTTACT   600
TGGCCTGGGT TTGCACCCAA TTTCTGGACA TTTTATAGCT GAGCATTACA TGTTCTTAAA   660
GGGTCATGAA ACTTACTCAT ATTATGGGCC TCTGAATTTA CTTACCTTCA ATGTGGGTTA   720
TCATAATGAA CATCATGATT TCCCCAACAT TCCTGGAAAA AGTCTTCCAC TGGTGAGGAA   780
AATAGCAGCT GAATACTATG ACAACCTCCC TCACTACAAT TCCTGGATAA AAGTACTGTA   840
```

TGATTTTGTG ATGGATGATA CAATAAGTCC CTACTCAAGA ATGAAGAGGC ACCAAAAAGG 900

AGAGATGGTG CTGGAGTAAA TATCATTAGT GCCAAAGGGA TTCTTCTCCA AAACTTTAGA 960

TGATAAAATG GAATTTTTGC ATTATTAAAC TTGAGACCAG TGATGCTCAG AAGCTCCCCT 1020

GGCACAATTT CAGAGTAAGA GCTCGGTGAT ACCAAGAAGT GAATCTGGCT TTTAAACAGT 1080

CAGCCTGACT CTGTACTGCT CAGTTTCACT CACAGGAAAC TTGTGACTTG TGTATTATCG 1140

TCATTGAGGA TGTTTCACTC ATGTCTGTCA TTTTATAAGC ATATCATTTA AAAAGCTTCT 1200

AAAAAGCTAT TTCGCCAGG 1219

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 655 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Edited Contig 2153526)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TTACCTTCTA CGTCCGCTTC TTCCTCACTT ATGTGCCACT ATTGGGGCTG AAAGCTTCCT 60

GGGCCTTTTC TTCATAGTCA GGTTCCTGGA AAGCAACTGG TTTGTGTGGG TGACACAGAT 120

GAACCATATT CCCATGCACA TTGATCATGA CCGGAACATG GACTGGGTTT CCACCCAGCT 180

CCAGGCCACA TGCAATGTCC ACAAGTCTGC CTTCAATGAC TGGTTCAGTG GACACCTCAA 240

CTTCCAGATT GAGCACCATC TTTTTCCCAC GATGCCTCGA CACAATTACC ACAAAGTGGC 300

TCCCCTGGTG CAGTCCTTGT GTGCCAAGCA TGGCATAGAG TACCAGTCCA AGCCCCTGCT 360

GTCAGCCTTC GCCGACATCA TCCACTCACT AAAGGAGTCA GGGCAGCTCT GGCTAGATGC 420

CTATCTTCAC CAATAACAAC AGCCACCCTG CCCAGTCTGG AAGAAGAGGA GGAAGACTCT 480

GGAGCCAAGG CAGAGGGGAG CTTGAGGGAC AATGCCACTA TAGTTTAATA CTCAGAGGGG 540

GTTGGGTTTG GGGACATAAA GCCTCTGACT CAAACTCCTC CCTTTTATCT TCTAGCCACA 600

GTTCTAAGAC CCAAAGTGGG GGGTGGACAC AGAAGTCCCT AGGAGGGAAG GAGCT 655

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 304 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Edited Contig 3506132)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GTCTTTTACT TTGGCAATGG CTGGATTCCT ACCCTCATCA CGGCCTTTGT CCTTGCTACC 60

TCTCAGGCCC AAGCTGGATG GCTGCAACAT GATTATGGCC ACCTGTCTGT CTACAGAAAA 120

CCCAAGTGGA ACCACCTTGT CCACAAATTC GTCATTGGCC ACTTAAAGGG TGCCTCTGCC 180

AACTGGTGGA ATCATCGCCA CTTCCAGCAC CACGCCAAGC CTAACATCTT CCACAAGGAT 240

CCCGATGTGA ACATGCTGCA CGTGTTTGTT CTGGGCGAAT GGCAGCCCAT CGAGTACGGC 300

AAGA 304

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 918 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Edited Contig 3854933)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CAGGGACCTA CCCCGCGCTA CTTCACCTGG GACGAGGTGG CCCAGCGCTC AGGGTGCGAG 60

GAGCGGTGGC TAGTGATCGA CCGTAAGGTG TACAACATCA GCGAGTTCAC CCGCCGGCAT 120

CCAGGGGGCT CCCGGGTCAT CAGCCACTAC GCCGGGCAGG ATGCCACGGA TCCCTTTGTG 180

GCCTTCCACA TCAACAAGGG CCTTGTGAAG AAGTATATGA ACTCTCTCCT GATTGGAGAA 240

CTGTCTCCAG AGCAGCCCAG CTTTGAGCCC ACCAAGAATA AAGAGCTGAC AGATGAGTTC 300

CGGGAGCTGC GGGCCACAGT GGAGCGGATG GGGCTCATGA AGGCCAACCA TGTCTTCTTC 360

CTGCTGTACC TGCTGCACAT CTTGCTGCTG GATGGTGCAG CCTGGCTCAC CCTTTGGGTC 420

TTTGGGACGT CCTTTTTGCC CTTCCTCCTC TGTGCGGTGC TGCTCAGTGC AGTTCAGGCC 480

CAGGCTGGCT GGCTGCAGCA TGACTTTGGG CACCTGTCGG TCTTCAGCAC CTCAAAGTGG 540

AACCATCTGC TACATCATTT TGTGATTGGC CACCTGAAGG GGGCCCCCGC CAGTTGGTGG 600

AACCACATGC ACTTCCAGCA CCATGCCAAG CCCAACTGCT TCCGCAAAGA CCCAGACATC 660

AACATGCATC CCTTCTTCTT TGCCTTGGGG AAGATCCTCT CTGTGGAGCT TGGGAAACAG 720

AAGAAAAAAT ATATGCCGTA CAACCACCAG CACARATACT TCTTCCTAAT TGGGCCCCCA 780

GCCTTGCTGC CTCTCTACTT CCAGTGGTAT ATTTTCTATT TTGTTATCCA GCGAAAGAAG 840

TGGGTGGACT TGGCCTGGAT CAGCAAACAG GAATACGATG AAGCCGGGCT TCCATTGTCC 900

ACCGCAAATG CTTCTAAA 918

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1686 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Edited Contig 2511785)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```
GCCACTTAAA GGGTGCCTCT GCCAACTGGT GGAATCATCG CCACTTCCAG CACCACGCCA   60
AGCCTAACAT CTTCCACAAG GATCCCGATG TGAACATGCT GCACGTGTTT GTTCTGGGCG  120
AATGGCAGCC CATCGAGTAC GGCAAGAAGA AGCTGAAATA CCTGCCCTAC AATCACCAGC  180
ACGAATACTT CTTCCTGATT GGGCCGCCGC TGCTCATCCC CATGTATTTC CAGTACCAGA  240
TCATCATGAC CATGATCGTC CATAAGAACT GGGTGGACCT GGCCTGGGCC GTCAGCTACT  300
ACATCCGGTT CTTCATCACC TACATCCCTT TCTACGGCAT CCTGGGAGCC CTCCTTTTCC  360
TCAACTTCAT CAGGTTCCTG GAGAGCCACT GGTTTGTGTG GGTCACACAG ATGAATCACA  420
TCGTCATGGA GATTGACCAG GAGGCCTACC GTGACTGGTT CAGTAGCCAG CTGACAGCCA  480
CCTGCAACGT GGAGCAGTCC TTCTTCAACG ACTGGTTCAG TGGACACCTT AACTTCCAGA  540
TTGAGCACCA CCTCTTCCCC ACCATGCCCC GGCACAACTT ACACAAGATC GCCCCGCTGG  600
TGAAGTCTCT ATGTGCCAAG CATGGCATTG AATACCAGGA GAAGCCGCTA CTGAGGGCCC  660
TGCTGGACAT CATCAGGTCC CTGAAGAAGT CTGGGAAGCT GTGGCTGGAC GCCTACCTTC  720
ACAAATGAAG CCACAGCCCC CGGGACACCG TGGGGAAGGG GTGCAGGTGG GGTGATGGCC  780
AGAGGAATGA TGGGCTTTTG TTCTGAGGGG TGTCCGAGAG GCTGGTGTAT GCACTGCTCA  840
CGGACCCCAT GTTGGATCTT TCTCCCTTTC TCCTCTCCTT TTTCTCTTCA CATCTCCCCC  900
ATAGCACCCT GCCCTCATGG GACCTGCCCT CCCTCAGCCG TCAGCCATCA GCCATGGCCC  960
TCCCAGTGCC TCCTAGCCCC TTCTTCCAAG GAGCAGAGAG GTGGCCACCG GGGGTGGCTC 1020
TGTCCTACCT CCACTCTCTG CCCCTAAAGA TGGGAGGAGA CCAGCGGTCC ATGGGTCTGG 1080
CCTGTGAGTC TCCCCTTGCA GCCTGGTCAC TAGGCATCAC CCCCGCTTTG GTTCTTCAGA 1140
TGCTCTTGGG GTTCATAGGG GCAGGTCCTA GTCGGGCAGG GCCCCTGACC CTCCCGGCCT 1200
GGCTTCACTC TCCCTGACGG CTGCCATTGG TCCACCCTTT CATAGAGAGG CCTGCTTTGT 1260
TACAAAGCTC GGGTCTCCCT CCTGCAGCTC GGTTAAGTAC CCGAGGCCTC TCTTAAGATG 1320
TCCAGGGCCC CAGGCCCGCG GGCACAGCCA GCCCAAACCT TGGGCCCTGG AAGAGTCCTC 1380
CACCCCATCA CTAGAGTGCT CTGACCCTGG GCTTTCACGG GCCCCATTCC ACCGCCTCCC 1440
CAACTTGAGC CTGTGACCTT GGGACCAAAG GGGGAGTCCC TCGTCTCTTG TGACTCAGCA 1500
GAGGCAGTGG CCACGTTCAG GGAGGGGCCG GCTGGCCTGG AGGCTCAGCC CACCCTCCAG 1560
CTTTTCCTCA GGGTGTCCTG AGGTCCAAGA TTCTGGAGCA ATCTGACCCT TCTCCAAAGG 1620
CTCTGTTATC AGCTGGGCAG TGCCAGCCAA TCCCTGGCCA TTTGGCCCCA GGGGACGTGG 1680
GCCCTG                                                            1686
```

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1843 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Contig 2535)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
GTCTTTTACT TTGGCAATGG CTGGATTCCT ACCCTCATCA CGGCCTTTGT CCTTGCTACC   60
TCTCAGGCCC AAGCTGGATG GCTGCAACAT GATTATGGCC ACCTGTCTGT CTACAGAAAA  120
CCCAAGTGGA ACCACCTTGT CCACAAATTC GTCATTGGCC ACTTAAAGGG TGCCTCTGCC  180
AACTGGTGGA ATCATCGCCA CTTCCAGCAC CACGCCAAGC CTAACATCTT CCACAAGGAT  240
CCCGATGTGA ACATGCTGCA CGTGTTTGTT CTGGGCGAAT GGCAGCCCAT CGAGTACGGC  300
AAGAAGAAGC TGAAATACCT GCCCTACAAT CACCAGCACG AATACTTCTT CCTGATTGGG  360
CCGCCGCTGC TCATCCCCAT GTATTTCCAG TACCAGATCA TCATGACCAT GATCGTCCAT  420
AAGAACTGGG TGGACCTGGC CTGGGCCGTC AGCTACTACA TCCGGTTCTT CATCACCTAC  480
ATCCCTTTCT ACGGCATCCT GGGAGCCCTC CTTTTCCTCA ACTTCATCAG GTTCCTGGAG  540
AGCCACTGGT TTGTGTGGGT CACACAGATG AATCACATCG TCATGGAGAT TGACCAGGAG  600
GCCTACCGTG ACTGGTTCAG TAGCCAGCTG ACAGCCACCT GCAACGTGGA GCAGTCCTTC  660
TTCAACGACT GGTTCAGTGG ACACCTTAAC TTCCAGATTG AGCACCACCT CTTCCCCACC  720
ATGCCCCGGC ACAACTTACA CAAGATCGCC CCGCTGGTGA AGTCTCTATG TGCCAAGCAT  780
GGCATTGAAT ACCAGGAGAA GCCGCTACTG AGGGCCCTGC TGGACATCAT CAGGTCCCTG  840
AAGAAGTCTG GGAAGCTGTG GCTGGACGCC TACCTTCACA AATGAAGCCA CAGCCCCCGG  900
GACACCGTGG GGAAGGGGTG CAGGTGGGGT GATGGCCAGA GGAATGATGG GCTTTTGTTC  960
TGAGGGGTGT CCGAGAGGCT GGTGTATGCA CTGCTCACGG ACCCCATGTT GGATCTTTCT 1020
CCCTTTCTCC TCTCCTTTTT CTCTTCACAT CTCCCCCATA GCACCCTGCC CTCATGGGAC 1080
CTGCCCTCCC TCAGCCGTCA GCCATCAGCC ATGGCCCTCC CAGTGCCTCC TAGCCCCTTC 1140
TTCCAAGGAG CAGAGAGGTG GCCACCGGGG GTGGCTCTGT CCTACCTCCA CTCTCTGCCC 1200
CTAAAGATGG GAGGAGACCA GCGGTCCATG GGTCTGGCCT GTGAGTCTCC CCTTGCAGCC 1260
TGGTCACTAG GCATCACCCC CGCTTTGGTT CTTCAGATGC TCTTGGGGTT CATAGGGGCA 1320
GGTCCTAGTC GGGCAGGGCC CCTGACCCTC CCGGCCTGGC TTCACTCTCC CTGACGGCTG 1380
CCATTGGTCC ACCCTTTCAT AGAGAGGCCT GCTTTGTTAC AAAGCTCGGG TCTCCCTCCT 1440
GCAGCTCGGT TAAGTACCCG AGGCCTCTCT TAAGATGTCC AGGGCCCCAG GCCCGCGGGC 1500
ACAGCCAGCC CAAACCTTGG GCCCTGGAAG AGTCCTCCAC CCCATCACTA GAGTGCTCTG 1560
ACCCTGGGCT TTCACGGGCC CCATTCCACC GCCTCCCCAA CTTGAGCCTG TGACCTTGGG 1620
ACCAAAGGGG GAGTCCCTCG TCTCTTGTGA CTCAGCAGAG GCAGTGGCCA CGTTCAGGGA 1680
```

GGGGCCGGCT GGCCTGGAGG CTCAGCCCAC CCTCCAGCTT TTCCTCAGGG TGTCCTGAGG 1740
TCCAAGATTC TGGAGCAATC TGACCCTTCT CCAAAGGCTC TGTTATCAGC TGGGCAGTGC 1800
CAGCCAATCC CTGGCCATTT GGCCCCAGGG GACGTGGGCC CTG 1843

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2257 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Edited Contig 253538a)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CAGGGACCTA CCCCGCGCTA CTTCACCTGG GACGAGGTGG CCCAGCGCTC AGGGTGCGAG 60
GAGCGGTGGC TAGTGATCGA CCGTAAGGTG TACAACATCA GCGAGTTCAC CCGCCGGCAT 120
CCAGGGGGCT CCCGGGTCAT CAGCCACTAC GCCGGGCAGG ATGCCACGGA TCCCTTTGTG 180
GCCTTCCACA TCAACAAGGG CCTTGTGAAG AAGTATATGA ACTCTCTCCT GATTGGAGAA 240
CTGTCTCCAG AGCAGCCCAG CTTTGAGCCC ACCAAGAATA AAGAGCTGAC AGATGAGTTC 300
CGGGAGCTGC GGGCCACAGT GGAGCGGATG GGGCTCATGA AGGCCAACCA TGTCTTCTTC 360
CTGCTGTACC TGCTGCACAT CTTGCTGCTG GATGGTGCAG CCTGGCTCAC CCTTTGGGTC 420
TTTGGGACGT CCTTTTTGCC CTTCCTCCTC TGTGCGGTGC TGCTCAGTGC AGTTCAGCAG 480
GCCCAAGCTG GATGGCTGCA ACATGATTAT GGCCACCTGT CTGTCTACAG AAAACCCAAG 540
TGGAACCACC TTGTCCACAA ATTCGTCATT GGCCACTTAA AGGGTGCCTC TGCCAACTGG 600
TGGAATCATC GCCACTTCCA GCACCACGCC AAGCCTAACA TCTTCCACAA GGATCCCGAT 660
GTGAACATGC TGCACGTGTT TGTTCTGGGC GAATGGCAGC CCATCGAGTA CGGCAAGAAG 720
AAGCTGAAAT ACCTGCCCTA CAATCACCAG CACGAATACT TCTTCCTGAT TGGGCCGCCG 780
CTGCTCATCC CCATGTATTT CCAGTACCAG ATCATCATGA CCATGATCGT CCATAAGAAC 840
TGGGTGGACC TGGCCTGGGC CGTCAGCTAC TACATCCGGT TCTTCATCAC CTACATCCCT 900
TTCTACGGCA TCCTGGGAGC CCTCCTTTTC CTCAACTTCA TCAGGTTCCT GGAGAGCCAC 960
TGGTTTGTGT GGGTCACACA GATGAATCAC ATCGTCATGG AGATTGACCA GGAGGCCTAC 1020
CGTGACTGGT TCAGTAGCCA GCTGACAGCC ACCTGCAACG TGGAGCAGTC CTTCTTCAAC 1080
GACTGGTTCA GTGGACACCT TAACTTCCAG ATTGAGCACC ACCTCTTCCC CACCATGCCC 1140
CGGCACAACT TACACAAGAT CGCCCCGCTG GTGAAGTCTC TATGTGCCAA GCATGGCATT 1200
GAATACCAGG AGAAGCCGCT ACTGAGGGCC CTGCTGGACA TCATCAGGTC CCTGAAGAAG 1260
TCTGGGAAGC TGTGGCTGGA CGCCTACCTT CACAAATGAA GCCACAGCCC CCGGGACACC 1320
GTGGGGAAGG GGTGCAGGTG GGGTGATGGC CAGAGGAATG ATGGGCTTTT GTTCTGAGGG 1380
GTGTCCGAGA GGCTGGTGTA TGCACTGCTC ACGGACCCCA TGTTGGATCT TTCTCCCTTT 1440

```
CTCCTCTCCT TTTTCTCTTC ACATCTCCCC CATAGCACCC TGCCCTCATG GGACCTGCCC  1500
TCCCTCAGCC GTCAGCCATC AGCCATGGCC CTCCCAGTGC CTCCTAGCCC CTTCTTCCAA  1560
GGAGCAGAGA GGTGGCCACC GGGGGTGGCT CTGTCCTACC TCCACTCTCT GCCCCTAAAG  1620
ATGGGAGGAG ACCAGCGGTC CATGGGTCTG GCCTGTGAGT CTCCCCTTGC AGCCTGGTCA  1680
CTAGGCATCA CCCCCGCTTT GGTTCTTCAG ATGCTCTTGG GGTTCATAGG GGCAGGTCCT  1740
AGTCGGGCAG GGCCCCTGAC CCTCCCGGCC TGGCTTCACT CTCCCTGACG GCTGCCATTG  1800
GTCCACCCTT TCATAGAGAG GCCTGCTTTG TTACAAAGCT CGGGTCTCCC TCCTGCAGCT  1860
CGGTTAAGTA CCCGAGGCCT CTCTTAAGAT GTCCAGGGCC CCAGGCCCGC GGGCACAGCC  1920
AGCCCAAACC TTGGGCCCTG GAAGAGTCCT CCACCCCATC ACTAGAGTGC TCTGACCCTG  1980
GGCTTTCACG GGCCCCATTC CACCGCCTCC CCAACTTGAG CCTGTGACCT TGGGACCAAA  2040
GGGGGAGTCC CTCGTCTCTT GTGACTCAGC AGAGGCAGTG GCCACGTTCA GGGAGGGGCC  2100
GGCTGGCCTG GAGGCTCAGC CCACCCTCCA GCTTTTCCTC AGGGTGTCCT GAGGTCCAAG  2160
ATTCTGGAGC AATCTGACCC TTCTCCAAAG GCTCTGTTAT CAGCTGGGCA GTGCCAGCCA  2220
ATCCCTGGCC ATTTGGCCCC AGGGGACGTG GGCCCTG                           2257
```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 411 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 2692004)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
His Ala Asp Arg Arg Arg Glu Ile Leu Ala Lys Tyr Pro Glu Ile
1               5                   10                  15
Lys Ser Leu Met Lys Pro Asp Pro Asn Leu Ile Trp Ile Ile Ile
                20                  25                  30
Met Met Val Leu Thr Gln Leu Gly Ala Phe Tyr Ile Val Lys Asp
                35                  40                  45
Leu Asp Trp Lys Trp Val Ile Phe Gly Ala Tyr Ala Phe Gly Ser
                50                  55                  60
Cys Ile Asn His Ser Met Thr Leu Ala Ile His Glu Ile Ala His
                65                  70                  75
Asn Ala Ala Phe Gly Asn Cys Lys Ala Met Trp Asn Arg Trp Phe
                80                  85                  90
Gly Met Phe Ala Asn Leu Pro Ile Gly Ile Pro Tyr Ser Ile Ser
                95                  100                 105
Phe Lys Arg Tyr His Met Asp His His Arg Tyr Leu Gly Ala Asp
                110                 115                 120
Gly Val Asp Val Asp Ile Pro Thr Asp Phe Glu Gly Trp Phe Phe
                125                 130                 135
Cys Thr Ala Phe Arg Lys Phe Ile Trp Val Ile Leu Gln Pro Leu
                140                 145                 150
Phe Tyr Ala Phe Arg Pro Leu Phe Ile Asn Pro Lys Pro Ile Thr
                155                 160                 165
Tyr Leu Glu Val Ile Asn Thr Val Ala Gln Val Thr Phe Asp Ile
```

```
                    170                      175                      180
Leu Ile Tyr Tyr Phe Leu Gly Ile Lys Ser Leu Val Tyr Met Leu
                185                      190                      195
Ala Ala Ser Leu Leu Gly Leu Gly Leu His Pro Ile Ser Gly His
                200                      205                      210
Phe Ile Ala Glu His Tyr Met Phe Leu Lys Gly His Glu Thr Tyr
                215                      220                      225
Ser Tyr Tyr Gly Pro Leu Asn Leu Leu Thr Phe Asn Val Gly Tyr
                230                      235                      240
His Asn Glu His His Asp Phe Pro Asn Ile Pro Gly Lys Ser Leu
                245                      250                      255
Pro Leu Val Arg Lys Ile Ala Ala Glu Tyr Tyr Asp Asn Leu Pro
                260                      265                      270
His Tyr Asn Ser Trp Ile Lys Val Leu Tyr Asp Phe Val Met Asp
                275                      280                      285
Asp Thr Ile Ser Pro Tyr Ser Arg Met Lys Arg His Gln Lys Gly
                290                      295                      300
Glu Met Val Leu Glu *** Ile Ser Leu Val Pro Lys Gly Phe Phe
                305                      310                      315
Ser Lys Thr Leu Asp Asp Lys Met Glu Phe Leu His Tyr *** Thr
                320                      325                      330
*** Asp Gln *** Cys Ser Glu Ala Pro Leu Ala Gln Phe Gln Ser
                335                      340                      345
Lys Ser Ser Val Ile Pro Arg Ser Glu Ser Gly Phe *** Thr Val
                350                      355                      360
Ser Leu Thr Leu Tyr Cys Ser Val Ser Leu Thr Gly Asn Leu ***
                365                      370                      375
Leu Val Tyr Tyr Arg His *** Gly Cys Phe Thr His Val Cys His
                380                      385                      390
Phe Ile Ser Ile Ser Phe Lys Lys Leu Leu Lys Ser Tyr Phe Ala
                400                      405                      410
Arg
```

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 218 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 2153526)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
Tyr Leu Leu Arg Pro Leu Leu Pro His Leu Cys Ala Thr Ile Gly
1               5                        10                       15
Ala Glu Ser Phe Leu Gly Leu Phe Phe Ile Val Arg Phe Leu Glu
                20                       25                       30
Ser Asn Trp Phe Val Trp Val Thr Gln Met Asn His Ile Pro Met
                35                       40                       45
His Ile Asp His Asp Arg Asn Met Asp Trp Val Ser Thr Gln Leu
                50                       55                       60
Gln Ala Thr Cys Asn Val His Lys Ser Ala Phe Asn Asp Trp Phe
                65                       70                       75
Ser Gly His Leu Asn Phe Gln Ile Glu His His Leu Phe Pro Thr
                80                       85                       90
Met Pro Arg His Asn Tyr His Lys Val Ala Pro Leu Val Gln Ser
                95                       100                      105
Leu Cys Ala Lys His Gly Ile Glu Tyr Gln Ser Lys Pro Leu Leu
                110                      115                      120
Ser Ala Phe Ala Asp Ile Ile His Ser Leu Lys Glu Ser Gly Gln
                125                      130                      135
Leu Trp Leu Asp Ala Tyr Leu His Gln *** Gln Gln Pro Pro Cys
                140                      145                      150
```

```
Pro Val Trp Lys Lys Arg Arg Lys Thr Leu Glu Pro Arg Gln Arg
                155                 160                 165
Gly Ala *** Gly Thr Met Pro Leu *** Phe Asn Thr Gln Arg Gly
                170                 175                 180
Leu Gly Leu Gly Thr *** Ser Leu *** Leu Lys Leu Leu Pro Phe
                185                 190                 195
Ile Phe *** Pro Gln Phe *** Asp Pro Lys Trp Gly Val Asp Thr
                200                 205                 210
Glu Val Pro Arg Arg Glu Gly Ala
                215
```

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 3506132)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```
Val Phe Tyr Phe Gly Asn Gly Trp Ile Pro Thr Leu Ile Thr Ala
1               5                   10                  15
Phe Val Leu Ala Thr Ser Gln Ala Gln Ala Gly Trp Leu Gln His
                20                  25                  30
Asp Tyr Gly His Leu Ser Val Tyr Arg Lys Pro Lys Trp Asn His
                35                  40                  45
Leu Val His Lys Phe Val Ile Gly His Leu Lys Gly Ala Ser Ala
                50                  55                  60
Asn Trp Trp Asn His Arg His Phe Gln His His Ala Lys Pro Asn
                65                  70                  75
Leu Gly Glu Trp Gln Pro Ile Glu Tyr Gly Lys Xxx
                80                  85
```

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 306 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 3854933)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

```
Gln Gly Pro Thr Pro Arg Tyr Phe Thr Trp Asp Glu Val Ala Gln
1               5                   10                  15
Arg Ser Gly Cys Glu Glu Arg Trp Leu Val Ile Asp Arg Lys Val
                20                  25                  30
Tyr Asn Ile Ser Glu Phe Thr Arg Arg His Pro Gly Gly Ser Arg
                35                  40                  45
Val Ile Ser His Tyr Ala Gly Gln Asp Ala Thr Asp Pro Phe Val
                50                  55                  60
Ala Phe His Ile Asn Lys Gly Leu Val Lys Lys Tyr Met Asn Ser
                65                  70                  75
Leu Leu Ile Gly Glu Leu Ser Pro Glu Gln Pro Ser Phe Glu Pro
```

```
                    80                       85                       90
Thr Lys Asn Lys Glu Leu Thr Asp Glu Phe Arg Glu Leu Arg Ala
                    95                      100                      105
Thr Val Glu Arg Met Gly Leu Met Lys Ala Asn His Val Phe Phe
                   110                      115                      120
Leu Leu Tyr Leu Leu His Ile Leu Leu Leu Asp Gly Ala Ala Trp
                   125                      130                      135
Leu Thr Leu Trp Val Phe Gly Thr Ser Phe Leu Pro Phe Leu Leu
                   140                      145                      150
Cys Ala Val Leu Leu Ser Ala Val Gln Ala Gln Ala Gly Trp Leu
                   155                      160                      165
Gln His Asp Phe Gly His Leu Ser Val Phe Ser Thr Ser Lys Trp
                   170                      175                      180
Asn His Leu Leu His His Phe Val Ile Gly His Leu Lys Gly Ala
                   185                      190                      195
Pro Ala Ser Trp Trp Asn His Met His Phe Gln His His Ala Lys
                   200                      205                      210
Pro Asn Cys Phe Arg Lys Asp Pro Asp Ile Asn Met His Pro Phe
                   215                      220                      225
Phe Phe Ala Leu Gly Lys Ile Leu Ser Val Glu Leu Gly Lys Gln
                   230                      235                      240
Lys Lys Lys Tyr Met Pro Tyr Asn His Gln His Xxx Tyr Phe Phe
                   245                      250                      255
Leu Ile Gly Pro Pro Ala Leu Leu Pro Leu Tyr Phe Gln Trp Tyr
                   260                      265                      270
Ile Phe Tyr Phe Val Ile Gln Arg Lys Lys Trp Val Asp Leu Ala
                   275                      280                      285
Trp Ile Ser Lys Gln Glu Tyr Asp Glu Ala Gly Leu Pro Leu Ser
                   290                      295                      300
Thr Ala Asn Ala Ser Lys
                   305
```

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 566 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 2511785)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

```
His Leu Lys Gly Ala Ser Ala Asn Trp Trp Asn His Arg His Phe
1                   5                       10                       15
Gln His His Ala Lys Pro Asn Ile Phe His Lys Asp Pro Asp Val
                    20                      25                       30
Asn Met Leu His Val Phe Val Leu Gly Glu Trp Gln Pro Ile Glu
                    35                      40                       45
Tyr Gly Lys Lys Lys Leu Lys Tyr Leu Pro Tyr Asn His Gln His
                    50                      55                       60
Glu Tyr Phe Phe Leu Ile Gly Pro Pro Leu Leu Ile Pro Met Tyr
                    65                      70                       75
Phe Gln Tyr Gln Ile Ile Met Thr Met Ile Val His Lys Asn Trp
                    80                      85                       90
Val Asp Leu Ala Trp Ala Val Ser Tyr Tyr Ile Arg Phe Phe Ile
                    95                     100                      105
Thr Tyr Ile Pro Phe Tyr Gly Ile Leu Gly Ala Leu Leu Phe Leu
                   110                     115                      120
Asn Phe Ile Arg Phe Leu Glu Ser His Trp Phe Val Trp Val Thr
                   125                     130                      135
Gln Met Asn His Ile Val Met Glu Ile Asp Gln Glu Ala Tyr Arg
                   140                     145                      150
```

```
Asp Trp Phe Ser Ser Gln Leu Thr Ala Thr Cys Asn Val Glu Gln
                155                 160                 165
Ser Phe Phe Asn Asp Trp Phe Ser Gly His Leu Asn Phe Gln Ile
                170                 175                 180
Glu His His Leu Phe Pro Thr Met Pro Arg His Asn Leu His Lys
                185                 190                 195
Ile Ala Pro Leu Val Lys Ser Leu Cys Ala Lys His Gly Ile Glu
                200                 205                 210
Tyr Gln Glu Lys Pro Leu Leu Arg Ala Leu Leu Asp Ile Ile Arg
                215                 220                 225
Ser Leu Lys Lys Ser Gly Lys Leu Trp Leu Asp Ala Tyr Leu His
                230                 235                 240
Lys *** Ser His Ser Pro Arg Asp Thr Val Gly Lys Gly Cys Arg
                245                 250                 255
Trp Gly Asp Gly Gln Arg Asn Asp Gly Leu Leu Phe *** Gly Val
                260                 265                 270
Ser Glu Arg Leu Val Tyr Ala Leu Leu Thr Asp Pro Met Leu Asp
                275                 280                 285
Leu Ser Pro Phe Leu Leu Ser Phe Phe Ser Ser His Leu Pro His
                290                 295                 300
Ser Thr Leu Pro Ser Trp Asp Leu Pro Ser Leu Ser Arg Gln Pro
                305                 310                 315
Ser Ala Met Ala Leu Pro Val Pro Pro Ser Pro Phe Phe Gln Gly
                320                 325                 330
Ala Glu Arg Trp Pro Pro Gly Val Ala Leu Ser Tyr Leu His Ser
                335                 340                 345
Leu Pro Leu Lys Met Gly Gly Asp Gln Arg Ser Met Gly Leu Ala
                350                 355                 360
Cys Glu Ser Pro Leu Ala Ala Trp Ser Leu Gly Ile Thr Pro Ala
                365                 370                 375
Leu Val Leu Gln Met Leu Leu Gly Phe Ile Gly Ala Gly Pro Ser
                380                 385                 390
Arg Ala Gly Pro Leu Thr Leu Pro Ala Trp Leu His Ser Pro ***
                400                 405                 410
Arg Leu Pro Leu Val His Pro Phe Ile Glu Arg Pro Ala Leu Leu
                415                 420                 425
Gln Ser Ser Gly Leu Pro Pro Ala Ala Arg Leu Ser Thr Arg Gly
                430                 435                 440
Leu Ser *** Asp Val Gln Gly Pro Arg Pro Ala Gly Thr Ala Ser
                445                 450                 455
Pro Asn Leu Gly Pro Trp Lys Ser Pro Pro Pro His His *** Ser
                460                 465                 470
Ala Leu Thr Leu Gly Phe His Gly Pro His Ser Thr Ala Ser Pro
                475                 480                 485
Thr *** Ala Cys Asp Leu Gly Thr Lys Gly Gly Val Pro Arg Leu
                490                 495                 500
Leu *** Leu Ser Arg Gly Ser Gly His Val Gln Gly Gly Ala Gly
                505                 510                 515
Trp Pro Gly Gly Ser Ala His Pro Pro Ala Phe Pro Gln Gly Val
                520                 525                 530
Leu Arg Ser Lys Ile Leu Glu Gln Ser Asp Pro Ser Pro Lys Ala
                535                 540                 545
Leu Leu Ser Ala Gly Gln Cys Gln Pro Ile Pro Gly His Leu Ala
                550                 555                 560
Pro Gly Asp Val Gly Pro Xxx
                565
```

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 619 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 2535)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

```
Val Phe Tyr Phe Gly Asn Gly Trp Ile Pro Thr Leu Ile Thr Ala
1               5                   10                  15
Phe Val Leu Ala Thr Ser Gln Ala Gln Ala Gly Trp Leu Gln His
                20                  25                  30
Asp Tyr Gly His Leu Ser Val Tyr Arg Lys Pro Lys Trp Asn His
                35                  40                  45
Leu Val His Lys Phe Val Ile Gly His Leu Lys Gly Ala Ser Ala
                50                  55                  60
Asn Trp Trp Asn His Arg His Phe Gln His His Ala Lys Pro Asn
                65                  70                  75
Ile Phe His Lys Asp Pro Asp Val Asn Met Leu His Val Phe Val
                80                  85                  90
Leu Gly Glu Trp Gln Pro Ile Glu Tyr Gly Lys Lys Lys Leu Lys
                95                  100                 105
Tyr Leu Pro Tyr Asn His Gln His Glu Tyr Phe Phe Leu Ile Gly
                110                 115                 120
Pro Pro Leu Leu Ile Pro Met Tyr Phe Gln Tyr Gln Ile Ile Met
                125                 130                 135
Thr Met Ile Val His Lys Asn Trp Val Asp Leu Ala Trp Ala Val
                140                 145                 150
Ser Tyr Tyr Ile Arg Phe Phe Ile Thr Tyr Ile Pro Phe Tyr Gly
                155                 160                 165
Ile Leu Gly Ala Leu Leu Phe Leu Asn Phe Ile Arg Phe Leu Glu
                170                 175                 180
Ser His Trp Phe Val Trp Val Thr Gln Met Asn His Ile Val Met
                185                 190                 195
Glu Ile Asp Gln Glu Ala Tyr Arg Asp Trp Phe Ser Ser Gln Leu
                200                 205                 210
Thr Ala Thr Cys Asn Val Glu Gln Ser Phe Phe Asn Asp Trp Phe
                215                 220                 225
Ser Gly His Leu Asn Phe Gln Ile Glu His His Leu Phe Pro Thr
                230                 235                 240
Met Pro Arg His Asn Leu His Lys Ile Ala Pro Leu Val Lys Ser
                245                 250                 255
Leu Cys Ala Lys His Gly Ile Glu Tyr Gln Glu Lys Pro Leu Leu
                260                 265                 270
Arg Ala Leu Leu Asp Ile Ile Arg Ser Leu Lys Lys Ser Gly Lys
                275                 280                 285
Leu Trp Leu Asp Ala Tyr Leu His Lys *** Ser His Ser Pro Arg
                290                 295                 300
Asp Thr Val Gly Lys Gly Cys Arg Trp Gly Asp Gly Gln Arg Asn
                305                 310                 315
Asp Gly Leu Leu Phe *** Gly Val Ser Glu Arg Leu Val Tyr Ala
                320                 325                 330
Leu Leu Thr Asp Pro Met Leu Asp Leu Ser Pro Phe Leu Leu Ser
                335                 340                 345
Phe Phe Ser Ser His Leu Pro His Ser Thr Leu Pro Ser Trp Asp
                350                 355                 360
Leu Pro Ser Leu Ser Arg Gln Pro Ser Ala Met Ala Leu Pro Val
                365                 370                 375
Pro Pro Ser Pro Phe Phe Gln Gly Ala Glu Arg Trp Pro Pro Gly
                380                 385                 390
Val Ala Leu Ser Tyr Leu His Ser Leu Pro Leu Lys Met Gly Gly
                400                 405                 410
Asp Gln Arg Ser Met Gly Leu Ala Cys Glu Ser Pro Leu Ala Ala
                415                 420                 425
Trp Ser Leu Gly Ile Thr Pro Ala Leu Val Leu Gln Met Leu Leu
                430                 435                 440
Gly Phe Ile Gly Ala Gly Pro Ser Arg Ala Gly Pro Leu Thr Leu
                445                 450                 455
```

```
Pro Ala Trp Leu His Ser Pro *** Arg Leu Pro Leu Val His Pro
                460                 465                 470
Phe Ile Glu Arg Pro Ala Leu Leu Gln Ser Ser Gly Leu Pro Pro
                475                 480                 485
Ala Ala Arg Leu Ser Thr Arg Gly Leu Ser *** Asp Val Gln Gly
                490                 495                 500
Pro Arg Pro Ala Gly Thr Ala Ser Pro Asn Leu Gly Pro Trp Lys
                505                 510                 515
Ser Pro Pro Pro His His *** Ser Ala Leu Thr Leu Gly Phe His
                520                 525                 530
Gly Pro His Ser Thr Ala Ser Pro Thr *** Ala Cys Asp Leu Gly
                535                 540                 545
Thr Lys Gly Gly Val Pro Arg Leu Leu *** Leu Ser Arg Gly Ser
                550                 555                 560
Gly His Val Gln Gly Gly Ala Gly Trp Pro Gly Gly Ser Ala His
                565                 570                 575
Pro Pro Ala Phe Pro Gln Gly Val Leu Arg Ser Lys Ile Leu Glu
                580                 585                 590
Gln Ser Asp Pro Ser Pro Lys Ala Leu Leu Ser Ala Gly Gln Cys
                595                 600                 605
Gln Pro Ile Pro Gly His Leu Ala Pro Gly Asp Val Gly Pro Xxx
                610                 615                 620
```

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 757 amino acids
  (B) TYPE: amino acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 253538a)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

```
Gln Gly Pro Thr Pro Arg Tyr Phe Thr Trp Asp Glu Val Ala Gln
1               5                   10                  15
Arg Ser Gly Cys Glu Glu Arg Trp Leu Val Ile Asp Arg Lys Val
                20                  25                  30
Tyr Asn Ile Ser Glu Phe Thr Arg Arg His Pro Gly Gly Ser Arg
                35                  40                  45
Val Ile Ser His Tyr Ala Gly Gln Asp Ala Thr Asp Pro Phe Val
                50                  55                  60
Ala Phe His Ile Asn Lys Gly Leu Val Lys Lys Tyr Met Asn Ser
                65                  70                  75
Leu Leu Ile Gly Glu Leu Ser Pro Glu Gln Pro Ser Phe Glu Pro
                80                  85                  90
Thr Lys Asn Lys Glu Leu Thr Asp Glu Phe Arg Glu Leu Arg Ala
                95                  100                 105
Thr Val Glu Arg Met Gly Leu Met Lys Ala Asn His Val Phe Phe
                110                 115                 120
Leu Leu Tyr Leu Leu His Ile Leu Leu Leu Asp Gly Ala Ala Trp
                125                 130                 135
Leu Thr Leu Trp Val Phe Gly Thr Ser Phe Leu Pro Phe Leu Leu
                140                 145                 150
Cys Ala Val Leu Leu Ser Ala Val Gln Gln Ala Gln Ala Gly Trp
                155                 160                 165
Leu Gln His Asp Tyr Gly His Leu Ser Val Tyr Arg Lys Pro Lys
                170                 175                 180
Trp Asn His Leu Val His Lys Phe Val Ile Gly His Leu Lys Gly
                185                 190                 195
Ala Ser Ala Asn Trp Trp Asn His Arg His Phe Gln His His Ala
                200                 205                 210
```

Lys Pro Asn Ile Phe His Lys Asp Pro Asp Val Asn Met Leu His
215 220 225
Val Phe Val Leu Gly Glu Trp Gln Pro Ile Glu Tyr Gly Lys Lys
230 235 240
Lys Leu Lys Tyr Leu Pro Tyr Asn His Gln His Glu Tyr Phe Phe
245 250 255
Leu Ile Gly Pro Pro Leu Leu Ile Pro Met Tyr Phe Gln Tyr Gln
260 265 270
Ile Ile Met Thr Met Ile Val His Lys Asn Trp Val Asp Leu Ala
275 280 285
Trp Ala Val Ser Tyr Tyr Ile Arg Phe Phe Ile Thr Tyr Ile Pro
290 295 300
Phe Tyr Gly Ile Leu Gly Ala Leu Leu Phe Leu Asn Phe Ile Arg
305 310 315
Phe Leu Glu Ser His Trp Phe Val Trp Val Thr Gln Met Asn His
320 325 330
Ile Val Met Glu Ile Asp Gln Glu Ala Tyr Arg Asp Trp Phe Ser
335 340 345
Ser Gln Leu Thr Ala Thr Cys Asn Val Glu Gln Ser Phe Phe Asn
350 355 360
Asp Trp Phe Ser Gly His Leu Asn Phe Gln Ile Glu His His Leu
365 370 375
Phe Pro Thr Met Pro Arg His Asn Leu His Lys Ile Ala Pro Leu
380 385 390
Val Lys Ser Leu Cys Ala Lys His Gly Ile Glu Tyr Gln Glu Lys
400 405 410
Pro Leu Leu Arg Ala Leu Leu Asp Ile Ile Arg Ser Leu Lys Lys
415 420 425
Ser Gly Lys Leu Trp Leu Asp Ala Tyr Leu His Lys *** Ser His
430 435 440
Ser Pro Arg Asp Thr Val Gly Lys Gly Cys Arg Trp Gly Asp Gly
445 450 455
Gln Arg Asn Asp Gly Leu Leu Phe *** Gly Val Ser Glu Arg Leu
460 465 470
Val Tyr Ala Leu Leu Thr Asp Pro Met Leu Asp Leu Ser Pro Phe
475 480 485
Leu Leu Ser Phe Phe Ser Ser His Leu Pro His Ser Thr Leu Pro
490 495 500
Ser Trp Asp Leu Pro Ser Leu Ser Arg Gln Pro Ser Ala Met Ala
505 510 515
Leu Pro Val Pro Pro Ser Pro Phe Phe Gln Gly Ala Glu Arg Trp
520 525 530
Pro Pro Gly Val Ala Leu Ser Tyr Leu His Ser Leu Pro Leu Lys
535 540 545
Met Gly Gly Asp Gln Arg Ser Met Gly Leu Ala Cys Glu Ser Pro
550 555 560
Leu Ala Ala Trp Ser Leu Gly Ile Thr Pro Ala Leu Val Leu Gln
565 570 575
Met Leu Leu Gly Phe Ile Gly Ala Gly Pro Ser Arg Ala Gly Pro
580 585 590
Leu Thr Leu Pro Ala Trp Leu His Ser Pro *** Arg Leu Pro Leu
595 600 605
Val His Pro Phe Ile Glu Arg Pro Ala Leu Leu Gln Ser Ser Gly
610 615 620
Leu Pro Pro Ala Ala Arg Leu Ser Thr Arg Gly Leu Ser *** Asp
625 630 635
Val Gln Gly Pro Arg Pro Ala Gly Thr Ala Ser Pro Asn Leu Gly
640 645 650
Pro Trp Lys Ser Pro Pro Pro His His *** Ser Ala Leu Thr Leu
655 660 665
Gly Phe His Gly Pro His Ser Thr Ala Ser Pro Thr *** Ala Cys
670 675 680
Asp Leu Gly Thr Lys Gly Gly Val Pro Arg Leu Leu *** Leu Ser
685 690 695
Arg Gly Ser Gly His Val Gln Gly Gly Ala Gly Trp Pro Gly Gly
700 705 710

```
Ser Ala His Pro Pro Ala Phe Pro Gln Gly Val Leu Arg Ser Lys
                715                 720                 725
Ile Leu Glu Gln Ser Asp Pro Ser Pro Lys Ala Leu Leu Ser Ala
                730                 735                 740
Gly Gln Cys Gln Pro Ile Pro Gly His Leu Ala Pro Gly Asp Val
                745                 750                 755
Gly Pro Xxx
```

What is claimed is:

1. An isolated nucleic acid comprising:

a nucleotide sequence depicted in SEQ ID NO: 1 or SEQ ID NO: 3.

2. A polypeptide encoded by a nucleotide sequence according to claim 1.

3. A purified or isolated polypeptide comprising an amino acid sequence depicted in SEQ ID NO: 2 or SEQ ID NO: 4.

4. An isolated nucleic acid encoding a polypeptide having an amino acid sequence depicted in SEQ ID NO: 2 or SEQ ID NO: 4.

5. An isolated nucleic acid comprising a nucleotide sequence which encodes a polypeptide which desaturates a fatty acid molecule at carbon 6 or 12 from the carboxyl end of said polypeptide, wherein said nucleotide sequence has an average A/T content of less than about 60%.

6. The isolated nucleic acid according to Claim 5, wherein said nucleic acid is derived from a fungus.

7. The isolated nucleic acid according to Claim 6, wherein said fungus is of the genus *Mortierella*.

8. The isolated nucleic acid according to Claim 7, wherein said fungus is of the species *Mortierella alpina*.

9. An isolated nucleic acid, wherein the nucleotide sequence of said nucleic acid is depicted in SEQ ID NO: 1. or SEQ ID NO: 3.

10. An isolated or purified polypeptide which desaturates a fatty acid molecule at carbon 6 or 12 from the carboxyl end of said polypeptide, wherein said polypeptide is a eukaryotic polypeptide or is derived from a eukaryotic polypeptide.

11. The isolated or purified eukaryotic polypeptide according to Claim 10, wherein said eukaryotic polypeptide is derived from a fungus.

12. A nucleic acid comprising:

a fungal nucleotide sequence which is substantially identical to a sequence of at least 50 nucleotides in SEQ ID NO: 1 or SEQ ID NO: 3 or is complementary to a sequence of at least 50 nucleotides in SEQ ID NO: 1 or SEQ ID NO: 3.

13. An isolated nucleic acid having a nucleotide sequence with at least about 50% homology to SEQ ID NO: 1 or SEQ ID NO: 3.

14. An isolated nucleic acid having a nucleotide sequence with at least about 50% homology to sequence encoding an amino acid sequence depicted in SEQ ID NO: 2 or SEQ ID NO: 4.

15. The nucleic acid of claim 14, wherein said amino acid sequence depicted in SEQ ID NO: 2 is selected from the group consisting of amino acid residues 50-53, 39-43, 172-176, 204-213, and 390-402.

16. A nucleic acid construct comprising:

a nucleotide sequence depicted in a SEQ ID NO: 1 or SEQ ID NO: 3 linked to a heterologous nucleic acid.

17. A nucleic acid construct comprising:

a nucleotide sequence depicted in a SEQ ID NO: 1 or SEQ ID NO: 3 operably associated with an expression control sequence functional in a microbial cell.

18. The nucleic acid construct according to Claim 17, wherein said microbial cell is a yeast cell.

19. The nucleic acid construct according to Claim 17, wherein said nucleotide sequence is derived from a fungus.

20. The nucleic acid construct according to Claim 19, wherein said fungus is of the genus *Mortierella*.

21. The nucleic acid construct according to Claim 20, wherein said fungus is of the species *Mortierella alpina*.

22. A nucleic acid construct comprising:

a fungal nucleotide sequence which encodes a polypeptide comprising an amino acid sequence which corresponds to or is complementary to an amino acid sequence depicted in SEQ ID NO: 2 or SEQ ID NO: 4, wherein said nucleic acid is operably associated with an expression control sequence functional in a microbial cell, wherein said nucleotide sequence encodes a functionally active polypeptide which desaturates a fatty acid molecule at carbon 6 or 12 from the carboxyl end of a fatty acid molecule.

23. A nucleic acid construct comprising:

a nucleotide sequence having an A/T content of less than about 60% which encodes a functionally active Δ6-desaturase having an amino acid sequence which corresponds to or is complementary to all of or a portion of an amino acid sequence depicted in a SEQ ID NO: 2, wherein said nucleotide sequence is operably associated with a transcription control sequence functional in a yeast cell.

24. A nucleic acid construct comprising:

a fungal nucleotide sequence which encodes a functionally active Δ12-desaturase having an amino acid sequence which corresponds to or is complementary to all of or a portion of an amino acid sequence depicted in a SEQ ID NO: 4, wherein said nucleotide sequence is operably associated with a transcription control sequence functional in a yeast cell.

25. A recombinant yeast cell comprising:

a nucleic acid construct according to Claim 23 or Claim 24.

26. The recombinant yeast cell according to Claim 25, wherein said yeast cell is a *Saccharomyces* cell.

27. A recombinant yeast cell comprising:

at least one copy of a vector comprising a fungal nucleotide sequence which encodes a polypeptide which converts 18:2 fatty acids to 18:3 fatty acids or 18:3 fatty acids to 18:4 fatty acids, wherein said yeast cell or an ancestor of said yeast cell was transformed with said vector to produce said recombinant yeast cell, and wherein said nucleotide sequence is operably associated with an expression control sequence functional in said recombinant yeast cell.

28. The recombinant yeast cell according to claim 27, wherein said fungal nucleotide sequence is a *Mortierella* nucleotide sequence.

29. The recombinant yeast cell according to Claim 28, wherein said recombinant yeast cell is a *Saccharomyces* cell.

30. The microbial cell according to Claim 27, wherein said expression control sequence is provided in said expression vector.

31. A method for production of GLA in a yeast culture, said method comprising:

growing a yeast culture having a plurality of recombinant yeast cells, wherein said yeast cells or an ancestor of said yeast cells were transformed with a vector comprising fungal DNA encoding a polypeptide which converts LA to GLA, wherein said DNA is operably associated with an expression control sequence functional in said yeast cells, under conditions whereby said DNA is expressed, whereby GLA is produced from LA in said yeast culture.

32. The method according to Claim 31, wherein said fungal DNA is *Mortierella* DNA and said polypeptide is a Δ6 desaturase.

33. The method according to Claim 32, wherein *Mortierella* is of the species *Mortierella alpina*.

34. The method according to Claim 31, wherein said LA is exogenously supplied.

35. The method according to Claim 31, wherein said conditions are inducible.

36. A method for production of stearidonic acid in a yeast culture, said method comprising:

growing a yeast culture having a plurality of recombinant yeast cells, wherein said yeast cells or an ancestor of said yeast cells were transformed with a vector comprising fungal DNA encoding a polypeptide which converts α-linolenic acid to stearidonic acid, wherein said DNA is operably associated with an expression control sequence functional in said yeast cells, under conditions whereby said DNA is expressed, whereby stearidonic acid is produced from α-linolenic acid in said yeast culture.

37. The method according to Claim 36, wherein said fungal DNA is *Mortierella* DNA and said polypeptide is a Δ6 desaturase.

38. The method according to Claim 37, wherein *Mortierella* is of the species *Mortierella alpina*.

39. The method according to Claim 36, wherein said α-linolenic acid is exogenously supplied.

40. The method according to Claim 36, wherein said conditions are inducible.

41. A method for production of linoleic acid in a yeast culture, said method comprising:

growing a yeast culture having a plurality of recombinant yeast cells, wherein said yeast cells or an ancestor of said yeast cells were transformed with a vector comprising fungal DNA encoding a polypeptide which converts oleic acid to linoleic acid, wherein said DNA is operably associated with an expression control sequence functional in said yeast cells, under conditions whereby said DNA is expressed, whereby linoleic acid is produced from oleic acid in said yeast culture.

42. The method according to Claim 41, wherein said fungal DNA is *Mortierella* DNA and said polypeptide is a Δ12 desaturase.

43. The method according to Claim 42, wherein *Mortierella* is of the species *Mortierella alpina*.

44. The method according to Claim 41, wherein said conditions are inducible.

45. An isolated or purified polypeptide which desaturates a fatty acid molecule at carbon 12 from the carboxyl end of said polypeptide, wherein said polypeptide is a fungal polypeptide or is derived from a fungal polypeptide.

46. The isolated or purified polypeptide according to Claim 46, wherein said polypeptide is a *Mortierrella alpina* Δ12 desaturase.

47. An isolated or purified polypeptide which desaturates a fatty acid molecule at carbon 6 from the carboxyl end of said polypeptide, wherein said polypeptide is a fungal polypeptide or is derived from a fungal polypeptide.

48. The isolated or purified polypeptide according to Claim 48, wherein said polypeptide is a Δ6 desaturase.

49. An isolated nucleic acid encoding a polypeptide according to Claim 47 or Claim 49.

50. The nucleic acid construct according to Claim 23, wherein said portion of an amino acid sequence depicted in SEQ.ID. NO: 2 comprises amino acids 1 through 457.

51. A host cell comprising:

a nucleic acid construct according to any one of Claims 22 to 24.

52. A host cell comprising:

a vector which includes a nucleic acid which encodes a fatty acid desaturase derived from *Mortierella alpina*, wherein said desaturase has an amino acid sequence represented by SEQ ID NO:2, and wherein said nucleotide sequence is operably linked to a promoter.

53. The host cell according to Claim 52, wherein said host cell is a eukaryotic cell.

54. The host cell according to Claim 53, wherein said eukaryotic cell is selected from the group consisting of a mammalian cell, a plant cell, an insect cell, a fungal cell, an avian cell and an algal cell.

55. The host cell according to Claim 54, wherein said host cell is a fungal cell.

56. The host cell of Claim 21, wherein said promoter is exogenously supplied to said host cell.

57. A method for production of stearidonic acid in a eukaryotic cell culture, said method comprising:

growing a eukaryotic cell culture having a plurality of recombinant eukaryotic cells, wherein said recombinant eukaryotic cells or ancestors of said recombinant eukaryotic cells were transformed with a vector comprising fungal DNA encoding a polypeptide which converts α-linolenic acid to stearidonic acid, wherein said DNA is operably associated with an expression control sequence functional in said recombinant eukaryotic cells, under conditions whereby said DNA is expressed, whereby stearidonic acid is produced from α-linolenic acid in said eukaryotic cell culture.

58. A method for production of linoleic acid in a eukaryotic cell culture, said method comprising:

growing a eukaryotic cell culture having a plurality of recombinant eukaruyotic cells, wherein said recombinant eukaryotic cells or ancestors of said recombinant eukaryotic cells were transformed with a vector comprising fungal DNA encoding a polypeptide which converts oleic acid to linoleic acid, wherein said DNA is operably associated with an expression control sequence functional in said recombinant eukaryotic cells, under conditions whereby said DNA is expressed, whereby linoleic acid is produced from oleic acid in said eukaryotic cell culture.

59. The method according to Claim 57 or Claim 58, wherein said eukaryotic cells are selected from the group consisting of mammalian cells, plant cells, insect cells, fungal cells, avian cells and algal cells.

60. The method according to Claim 59, wherein said fungal cells are yeast cells of the genus *Saccharomyces*.

61. A recombinant yeast cell comprising:

(1) at least one nucleic acid construct according to Claim 23 or 24; or

(2) at least one nucleic acid construct according to Claim 23 and at least one nucleic acid construct according to Claim 24.

62. A recombinant yeast cell comprising:

at least one nucleic acid construct comprising a nucleotide sequence which encodes a functionally active Δ6 desaturase having an amino acid sequence which corresponds to or is complementary to all or a portion of an amino acid sequence depicted in SEQ ID NO: 2, and at least one nucleic acid construct comprising a nucleotide sequence which encodes a functionally active Δ12 desaturase having an amino acid sequence which corresponds to or is complementary to all or a portion of an amino acid sequence depicted in SEQ ID NO: 4, wherein said nucleic acid constructs are operably associated with transcription control sequences functional in a yeast cell.

63. A method of making GLA, said method comprising:

growing a recombinant yeast cell according to Claim 62 under conditions whereby said nucleotide sequences are expressed , whereby GLA is produced in said yeast cell.

64. A method of making GLA, said method comprising:

growing a recombinant yeast cell according to Claim 61 under conditions whereby the nucleotide sequences in said nucleic acid constructs are expressed , whereby GLA is produced in said yeast cell.

65. A method for obtaining altered long chain polyunsaturated fatty acid biosynthesis comprising the steps of:

growing a plant having cells which contain one or more transgenes, derived from a fungus or algae, which encodes a transgene expression product which desaturates a fatty acid molecule at a carbon selected from the group consisting of carbon 6 and carbon 12 from the carboxyl end of said fatty acid molecule, wherein said one or more trangenes is operably associated with an expression control sequence, under conditions whereby said one or more transgenes is expressed, whereby long chain polyunsaturated fatty acid biosynthesis in said cells is altered.

66. The method according to claim 65, wherein said long chain polyunsaturated fatty acid is selected from the group consisting of 18:1ω9, LA, GLA, SDA and ALA.

67. A microbial oil or fraction thereof produced according to the method of claim 65.

68. A method of treating or preventing malnutrition comprising administering said microbial oil of claim 67 to a patient in need of said treatment or prevention in an amount sufficient to effect said treatment or prevention.

69. A pharmaceutical composition comprising said microbial oil or fraction of claim 67 and a pharmaceutically acceptable carrier.

70. The pharmaceutical composition of claim 69, wherein said pharmaceutical composition is in the form of a solid or a liquid.

71. The pharmaceutical composition of claim 70, wherein said pharmaceutical composition is in a capsule or tablet form.

72. The pharmaceutical composition of claim 69 further comprising at least one nutrient selected from the group consisting of a vitamin, a mineral, a carbohydrate, a sugar, an amino acid, a free fatty acid, a phospholipid, an antioxidant, and a phenolic compound.

73. A nutritional formula comprising said microbial oil or fraction thereof of claim 67.

74. The nutritional formula of claim 73, wherein said nutritional formula is selected from the group consisting of an infant formula, a dietary supplement, and a dietary substitute.

75. The nutritional formula of claim 74, wherein said infant formula, dietary supplement or dietary supplement is in the form of a liquid or a solid.

76. An infant formula comprising said microbial oil or fraction thereof of claim 67.

77. The infant formula of claim 76 further comprising at least one macronutrient selected from the group consisting of coconut oil, soy oil, canola oil, mono- and diglycerides, glucose, edible lactose, electrodialysed whey, electrodialysed skim milk, milk whey, soy protein, and other protein hydrolysates.

78. The infant formula of claim 77 further comprising at least one vitamin selected from the group consisting of Vitamins A, C, D, E, and B complex; and at least one mineral selected from the group consisting of calcium, magnesium, zinc, manganese, sodium, potassium, phosphorus, copper, chloride, iodine, selenium, and iron.

79. A dietary supplement comprising said microbial oil or fraction thereof of claim 67.

80. The dietary supplement of claim 79 further comprising at least one macronutrient selected from the group consisting of coconut oil, soy oil, canola oil, mono- and diglycerides, glucose, edible lactose, electrodialysed whey, electrodialysed skim milk, milk whey, soy protein, and other protein hydrolysates.

81. The dietary supplement of claim 80 further comprising at least one vitamin selected from the group consisting of Vitamins A, C, D, E, and B complex; and at least one mineral selected from the group consisting of calcium, magnesium, zinc, manganese, sodium, potassium, phosphorus, copper, chloride, iodine, selenium, and iron.

82. The dietary supplement of claim 79 or claim 81, wherein said dietary supplement is administered to a human or an animal.

83. A dietary substitute comprising said microbial oil or fraction thereof of claim 67.

84. The dietary substitute of claim 83 further comprising at least one macronutrient selected from the group consisting of coconut oil, soy oil, canola oil, mono- and diglycerides, glucose, edible lactose, electrodialysed whey, electrodialysed skim milk, milk whey, soy protein, and other protein hydrolysates.

85. The dietary substitute of claim 84 further comprising at least one vitamin selected from the group consisting of Vitamins A, C, D, E, and B complex; and at least one mineral selected from the group c nsisting of calcium, magnesium,

zinc, manganese, sodium, potassium, phosphorus, copper, chloride, iodine, selenium, and iron.

86. The dietary substitute of claim 83 or claim 85, wherein said dietary substitute is administered to a human or animal.

87. A method of treating a patient having a condition caused by insuffient intake or production of polyunsaturated fatty acids comprising administering to said patient said dietary substitute of claim 83 or said dietary supplement of claim 79 in an amount sufficient to effect said treatment.

88. The method of claim 87, wherein said dietary substitute or said dietary supplement is administered enterally or parenterally.

89. A cosmetic comprising said microbial oil or fraction thereof of claim 67.

90. The cosmetic of claim 88, wherein said cosmetic is applied topically.

91. The pharmaceutical composition of claim 69, wherein said pharmaceutical composition is administered to a human or an animal.

92. An animal feed comprising said microbial oil or fraction thereof of claim 67.

93. The method of claim 20 wherein said fungus is *Mortierella species*.

94. The method of claim 93 wherein said fungus is *Mortierella alpina*.

95. An isolated peptide sequence selected from the group consisting of SEQ ID NO:34 - SEQ ID NO:40.

96. An isolated peptide sequence selected from the group consisting of SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:25 and SEQ ID NO:26.

97. A method for production of gamma-linolenic acid in a eukaryotic cell culture, said method comprising:

growing a eukaryotic cell culture having a plurality of recombinant eukaryotic cells, wherein said recombinant eukaryotic cells or ancestors of said recombinant eukaryotic cells were transformed with a vector comprising fungal DNA encoding a polypeptide which converts linoleic acid to gamma-linolenic acid, wherein said DNA is operably associated with an expression control sequence functional in said recombinant eukaryotic cells, under conditions whereby said DNA is expressed, whereby gamma-linolenic acid is produced from linoleic acid in said eukaryotic cell culture.

98. The method according to Claim 97 wherein said eukaryotic cells are selected from the group consisting of mammalian cells, plant cells, insect cells, fungal cells, avian cells and algal cells.

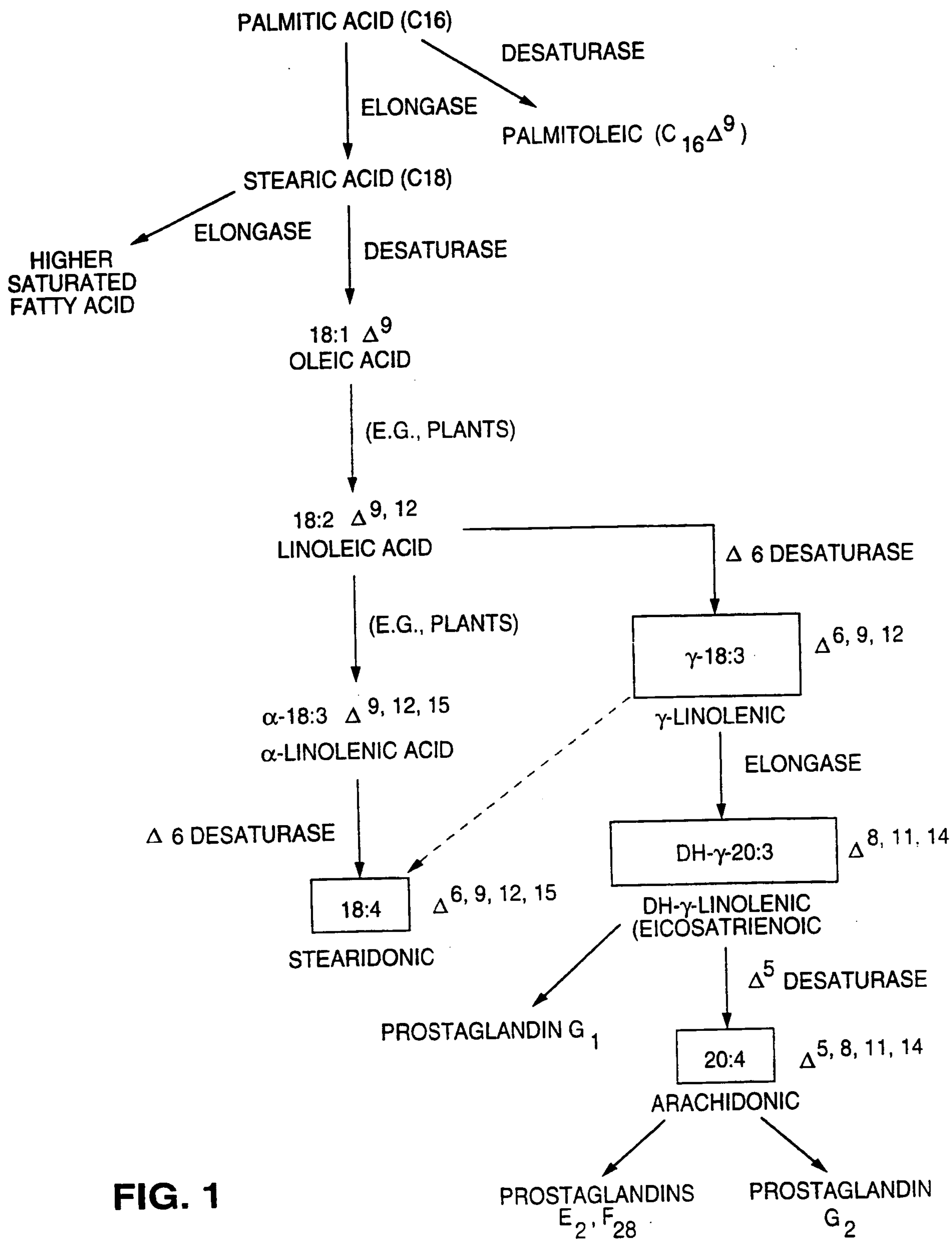

PALMITIC ACID (C16)
DESATURASE
ELONGASE
PALMITOLEIC ($C_{16}\Delta^9$)
STEARIC ACID (C18)
ELONGASE
DESATURASE
HIGHER SATURATED FATTY ACID
18:1 Δ^9
OLEIC ACID
(E.G., PLANTS)
18:2 $\Delta^{9, 12}$
LINOLEIC ACID
Δ 6 DESATURASE
(E.G., PLANTS)
γ-18:3
$\Delta^{6, 9, 12}$
γ-LINOLENIC
α-18:3 $\Delta^{9, 12, 15}$
α-LINOLENIC ACID
ELONGASE
Δ 6 DESATURASE
DH-γ-20:3
$\Delta^{8, 11, 14}$
18:4
$\Delta^{6, 9, 12, 15}$
STEARIDONIC
DH-γ-LINOLENIC
(EICOSATRIENOIC
Δ^5 DESATURASE
PROSTAGLANDIN G_1
20:4
$\Delta^{5, 8, 11, 14}$
ARACHIDONIC
PROSTAGLANDINS E_2, F_{28}
PROSTAGLANDIN G_2

FIG. 1

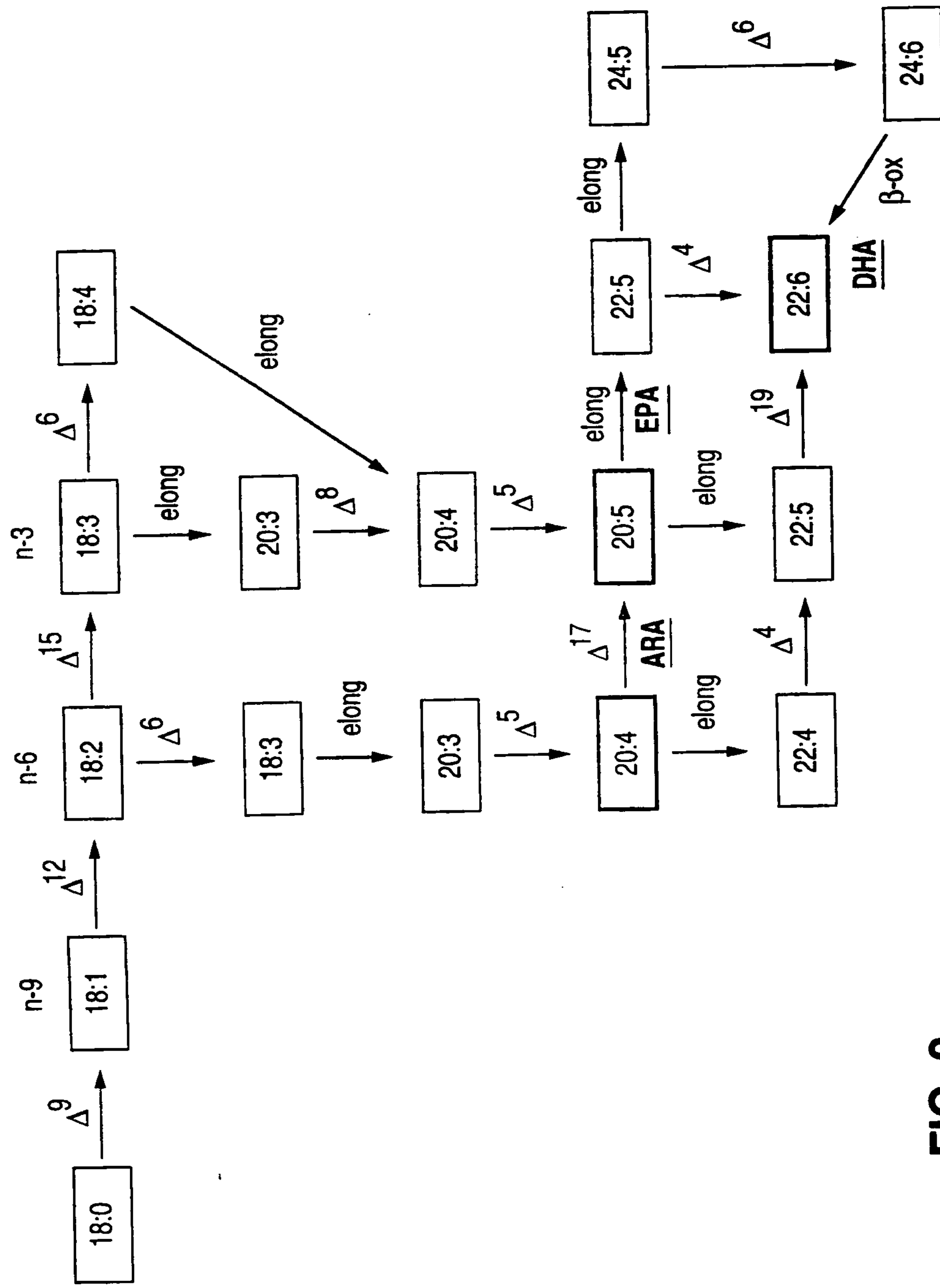

n-9
n-6
n-3
18:0
Δ^9
18:1
Δ^{12}
18:2
Δ^{15}
18:3
Δ^6
18:4
Δ^6
18:3
elong
20:3
Δ^5
20:4
elong
20:3
Δ^8
20:4
Δ^5
20:5
Δ^{17}
ARA
EPA
elong
22:4
Δ^4
22:5
Δ^{19}
22:6
DHA
elong
22:5
Δ^4
elong
24:5
Δ^6
24:6
β-ox

FIG. 2

```
                                                                     60
                                                                      *
CGACACTCCT TCCTTCTTCT CACCCGTCCT AGTCCCCTTC AACCCCCCTC TTTGACAAAG

ACAACAAACC ATG GCT GCT GCT CCC AGT GTG AGG ACG TTT ACT CGG GCC GAG
           Met Ala Ala Ala Pro Ser Val Arg Thr Phe Thr Arg Ala Glu

        120
         *
GTT TTG AAT GCC GAG GCT CTG AAT GAG GGC AAG AAG GAT GCC GAG GCA
Val Leu Asn Ala Glu Ala Leu Asn Glu Gly Lys Lys Asp Ala Glu Ala

                        180
                         *
CCC TTC TTG ATG ATC ATC GAC AAC AAG GTG TAC GAT GTC CGC GAG TTC
Pro Phe Leu Met Ile Ile Asp Asn Lys Val Tyr Asp Val Arg Glu Phe

                                        240
                                         *
GTC CCT GAT CAT CCC GGT GGA AGT GTG ATT CTC ACG CAC GTT GGC AAG
Val Pro Asp His Pro Gly Gly Ser Val Ile Leu Thr His Val Gly Lys

                                                        300
                                                         *
GAC GGC ACT GAC GTC TTT GAC ACT TTT CAC CCC GAG GCT GCT TGG GAG
Asp Gly Thr Asp Val Phe Asp Thr Phe His Pro Glu Ala Ala Trp Glu

ACT CTT GCC AAC TTT TAC GTT GGT GAT ATT GAC GAG AGC GAC CGC GAT
Thr Leu Ala Asn Phe Tyr Val Gly Asp Ile Asp Glu Ser Asp Arg Asp

        360
         *
ATC AAG AAT GAT GAC TTT GCG GCC GAG GTC CGC AAG CTG CGT ACC TTG
Ile Lys Asn Asp Asp Phe Ala A'la Glu Val Arg Lys Leu Arg Thr Leu
```

FIG. 3A

```
                                420
                                 *
TTC CAG TCT CTT GGT TAC TAC GAT TCT TCC AAG GCA TAC TAC GCC TTC
Phe Gln Ser Leu Gly Tyr Tyr Asp Ser Ser Lys Ala Tyr Tyr Ala Phe
                                            480
                                             *
AAG GTC TCG TTC AAC CTC TGC ATC TGG GGT TTG TCG ACG GTC ATT GTG
Lys Val Ser Phe Asn Leu Cys Ile Trp Gly Leu Ser Thr Val Ile Val
                                                        540
                                                         *
GCC AAG TGG GGC CAG ACC TCG ACC CTC GCC AAC GTG CTC TCG GCT GCG
Ala Lys Trp Gly Gln Thr Ser Thr Leu Ala Asn Val Leu Ser Ala Ala

CTT TTG GGT CTG TTC TGG CAG CAG TGC GGA TGG TTG GCT CAC GAC TTT
Leu Leu Gly Leu Phe Trp Gln Gln Cys Gly Trp Leu Ala His Asp Phe
        600
         *
TTG CAT CAC CAG GTC TTC CAG GAC CGT TTC TGG GGT GAT CTT TTC GGC
Leu His His Gln Val Phe Gln Asp Arg Phe Trp Gly Asp Leu Phe Gly
                            660
                             *
GCC TTC TTG GGA GGT GTC TGC CAG GGC TTC TCG TCC TCG TGG TGG AAG
Ala Phe Leu Gly Gly Val Cys Gln Gly Phe Ser Ser Ser Trp Trp Lys
                                        720
                                         *
GAC AAG CAC AAC ACT CAC CAC GCC GCC CCC AAC GTC CAC GGC GAG GAT
Asp Lys His Asn Thr His His Ala Ala Pro Asn Val His Gly Glu Asp
                                                            780
                                                             *
```

FIG. 3B

```
CCC GAC ATT GAC ACC CAC CCT CTG TTG ACC TGG AGT GAG CAT GCG TTG
Pro Asp Ile Asp Thr His Pro Leu Leu Thr Trp Ser Glu His Ala Leu

GAG ATG TTC TCG GAT GTC CCA GAT GAG GAG CTG ACC CGC ATG TGG TCG
Glu Met Phe Ser Asp Val Pro Asp Glu Glu Leu Thr Arg Met Trp Ser

        840
         *
CGT TTC ATG GTC CTG AAC CAG ACC TGG TTT TAC TTC CCC ATT CTC TCG
Arg Phe Met Val Leu Asn Gln Thr Trp Phe Tyr Phe Pro Ile Leu Ser

                        900
                         *
TTT GCC CGT CTC TCC TGG TGC CTC CAG TCC ATT CTC TTT GTG CTG CCT
Phe Ala Arg Leu Ser Trp Cys Leu Gln Ser Ile Leu Phe Val Leu Pro

                                        960
                                         *
AAC GGT CAG GCC CAC AAG CCC TCG GGC GCG CGT GTG CCC ATC TCG TTG
Asn Gly Gln Ala His Lys Pro Ser Gly Ala Arg Val Pro Ile Ser Leu

                                                       1020
                                                         *
GTC GAG CAG CTG TCG CTT GCG ATG CAC TGG ACC TGG TAC CTC GCC ACC
Val Glu Gln Leu Ser Leu Ala Met His Trp Thr Trp Tyr Leu Ala Thr

ATG TTC CTG TTC ATC AAG GAT CCC GTC AAC ATG CTG GTG TAC TTT TTG
Met Phe Leu Phe Ile Lys Asp Pro Val Asn Met Leu Val Tyr Phe Leu

       1080
         *
GTG TCG CAG GCG GTG TGC GGA AAC TTG TTG GCG ATC GTG TTC TCG CTC
Val Ser Gln Ala Val Cys Gly Asn Leu Leu Ala Ile Val Phe Ser Leu
```

FIG. 3C

```
                              1140
                                *
AAC CAC AAC GGT ATG CCT GTG ATC TCG AAG GAG GAG GCG GTC GAT ATG
Asn His Asn Gly Met Pro Val Ile Ser Lys Glu Glu Ala Val Asp Met
                                            1200
                                              *
GAT TTC TTC ACG AAG CAG ATC ATC ACG GGT CGT GAT GTC CAC CCG GGT
Asp Phe Phe Thr Lys Gln Ile Ile Thr Gly Arg Asp Val His Pro Gly
                                                        1260
                                                          *
CTA TTT GCC AAC TGG TTC ACG GGT GGA TTG AAC TAT CAG ATC GAG CAC
Leu Phe Ala Asn Trp Phe Thr Gly Gly Leu Asn Tyr Gln Ile Glu His

CAC TTG TTC CCT TCG ATG CCT CGC CAC AAC TTT TCA AAG ATC CAG CCT
His Leu Phe Pro Ser Met Pro Arg His Asn Phe Ser Lys Ile Gln Pro
        1320
          *
GCT GTC GAG ACC CTG TGC AAA AAG TAC AAT GTC CGA TAC CAC ACC ACC
Ala Val Glu Thr Leu Cys Lys Lys Tyr Asn Val Arg Tyr His Thr Thr
                          1380
                            *
GGT ATG ATC GAG GGA ACT GCA GAG GTC TTT AGC CGT CTG AAC GAG GTC
Gly Met Ile Glu Gly Thr Ala Glu Val Phe Ser Arg Leu Asn Glu Val
                                        1440
                                          *
TCC AAG GCT GCC TCC AAG ATG GGT AAG GCG CAG TAAAAAAAA AAACAAGGAC
Ser Lys Ala Ala Ser Lys Met Gly Lys Ala Gln
```

FIG. 3D

```
                                         1500
                                            *
GTTTTTTTTC GCCAGTGCCT GTGCCTGTGC CTGCTTCCCT TGTCAAGTCG AGCGTTTCTG
                                         1560
                                            *
GAAAGGATCG TTCAGTGCAG TATCATCATT CTCCTTTTAC CCCCCGCTCA TATCTCATTC

ATTTCTCTTA TTAAACAACT TGTTCCCCCC TTCACCG
```

FIG. 3E

```
Ma524      EVRKLRTLFQSLGYYDSSKAYYAFKVSFNLCIWGLSTVIVAKWGQTSTLANVLSAALLGL 90
ATTS4723   ---------------------VTLY-TLAFVAAMSLGVLYGVLACPSVXPHQIAAGLLGL 38
12-5       ------------------------------------GVLYGVLACTSVFAHQIAAALLGL 24
T42806     ------------------------------------------------------------ 4
W28140     -----GXX---------------------------------------------------- 1
R05219     --------------------------------C--------------------------- 2
W53753     ------------------------------------------------------------ 1

Ma524      FWQQCGWLAHDFLHHQVFQDRFWGDL-FGAFLGGVC-QGFSSSWWKDKHNTHHAAPNVHGE 119
ATTS4723   LWIQSAYIGXDSGHYVIMSNKSNNX-FAQLLSGNCLTGI  IAWWKWTHNAHHLACNSLDY 97
12-5       LWIQSAYIGHDSGHYVIMSNKSYNR-FAQLLSGNCLTGISIAWWKWTHNAHHLACNSLDY 83
T42806     ------------------------------------------------------------- 4
W28140     ------------------------------------------------------------- 1
R05219     ------------------------------------------------------------- 2
W53753     ------------------------------------------------------------- 1

Ma524      DPDIDTHPLLTWSEHALEMFSDVPDEELTRMWS----RFMVLNQTWFYFPILSFARLSW 174
ATTS4723   GPNLQHIP                                                    105
12-5       DPDLQHIPVFAVSTK---FFSSLTSRFYDRKLTFGPVARFLVSYQHFTYYPVMCFGRINL 140
T42806     ------------------------------------------------------------ 4
W28140     ------------------------------------------------------------ 1
R05219     ------------------------------------------------------------ 2
W53753     ------------------------------------------------------------ 1
```

FIG. 4A

```
Ma524     CLQSILFVLPNGQAHKPSGARVPISLVEQLSLAM-----HWTWYLATMFLFIKDPVNMLV 229
ATTS4723                                             W W                     105
12-5      FIQTFLLLFSKRE---------VPDRALNFAGILV-----FWTWF--PLLVSCLPNWPERF 185
T42806    --------------------------NFAGILV-----FFTVF--PLLVSCLPNWPERF 29
W28140    --------------------------PATEVGGLAWMIT-Y-RFFLTYVPLLGLKAFLG 33
R05219    ------------------------------------------F-S------------------- 2
W53753    -----------RHEAARGGTRLAYMLVCMQWTDL--LWAAS Y RFFLSYSPFYGATGTLL 48

Ma524     YFLVSQAVCGNLLAIVFSLNHNGMPVISKEEAVDMDFFTKQIITGRDVHPGLFANWFTGG 289
ATTS4723                                                               105
12-5      FFVFTSFTVTALQHIQFTLNHFAADVYV-GPPTGSDWFEKQAAGTIDISCRSYMDWFFGG 244
T42806    XFVFTGFTVTALQHIQFTLNHFAADVYV-GPPTGSDWFEKQAAGTIDISCRSYMDWFFGG 88
W28140    LFFIVRFLESNWFVWVTQMNH---IPMHIDHDRNMDWVSTQLQATCNVHKSAFNIDWFSGH 90
R05219    ----------------------------------------SPKSSPTRNMTPSPFIDWLWGG 23
W53753    LFVAVRVLESHWFVWITQMNH---IPKEIGHEKHRDWASSQLAATCNVEPSLF DWFSGH 105

Ma524     LNYQIEHHLFPSMPRHNFSKIQPAVETLCKKYNVRYHTTGMIEGTAEVESRLNEVSKAAS 349
ATTS4723                                                               105
12-5      LQFQLEHH                                                     252
T42806    LQFQLEHHLFPRLPRCHLRKVSPVGQRGFQRKXNLSX                        125
W28140    LNFQIEHHLFPTMPRHNYHXVAPLVQSLCAKHGIEYQSKPL                    131
R05219    LNYQIEHHLFPTMPRCNLNRCMKYVKEWCAENNLPYLVDDYFVGYNLNLQQLKNMAELVQ 83
W53753    LNFQIEHHLFPTMPRHNYRXVAPLVKAFCAKHGLHYEV                       143

Ma524     KMGKAQ                                                       355
ATTS4723                                                               105
12-5                                                                   252
T42806                                                                 125
W28140                                                                 131
R05219                                                                 87
W53753    --AKAA                                                       148
```

FIG. 4B

```
                                                                   60
                                                                    *
GTCCCCTGTC GCTGTCGGCA CACCCCATCC TCCCTCGCTC CCTCTGCGTT TGTCCTTGGC
                                                                  120
                                                                    *
CCACCGTCTC TCCTCCACCC TCCGAGACGA CTGCAACTGT AATCAGGAAC CGACAAATAC
                                                                      180
                                                                        *
ACGATTTCTT TTTACTCAGC ACCAACTCAA AATCCTCAAC CGCAACCCTT TTTCAGG ATG
                                                                   Met
GCA CCT CCC AAC ACT ATC GAT GCC GGT TTG ACC CAG CGT CAT ATC AGC
Ala Pro Pro Asn Thr Ile Asp Ala Gly Leu Thr Gln Arg His Ile Ser
            240
              *
ACC TCG GCC CCA AAC TCG GCC AAG CCT GCC TTC GAG CGC AAC TAC CAG
Thr Ser Ala Pro Asn Ser Ala Lys Pro Ala Phe Glu Arg Asn Tyr Gln
                            300
                              *
CTC CCC GAG TTC ACC ATC AAG GAG ATC CGA GAG TGC ATC CCT GCC CAC
Leu Pro Glu Phe Thr Ile Lys Glu Ile Arg Glu Cys Ile Pro Ala His
                                            360
                                              *
TGC TTT GAG CGC TCC GGT CTC CGT GGT CTC TGC CAC GTT GCC ATC GAT
Cys Phe Glu Arg Ser Gly Leu Arg Gly Leu Cys His Val Ala Ile Asp
                                                            420
                                                              *
CTG ACT TGG GCG TCG CTC TTG TTC CTG GCT GCG ACC CAG ATC GAC AAG
Leu Thr Trp Ala Ser Leu Leu Phe Leu Ala Ala Thr Gln Ile Asp Lys

TTT GAG AAT CCC TTG ATC CGC TAT TTG GCC TGG CCT GTT TAC TGG ATC
Phe Glu Asn Pro Leu Ile Arg Tyr Leu Ala Trp Pro Val Tyr Trp Ile
```

FIG. 5A

```
                  480
                    *
ATG CAG GGT ATT GTC TGC ACC GGT GTC TGG GTG CTG GCT CAC GAG TGT
Met Gln Gly Ile Val Cys Thr Gly Val Trp Val Leu Ala His Glu Cys
                                540
                                  *
GGT CAT CAG TCC TTC TCG ACC TCC AAG ACC CTC AAC AAC ACA GTT GGT
Gly His Gln Ser Phe Ser Thr Ser Lys Thr Leu Asn Asn Thr Val Gly
                                            600
                                              *
TGG ATC TTG CAC TCG ATG CTC TTG GTC CCC TAC CAC TCC TGG AGA ATC
Trp Ile Leu His Ser Met Leu Leu Val Pro Tyr His Ser Trp Arg Ile
                                                            660
                                                              *
TCG CAC TCG AAG CAC CAC AAG GCC ACT GGC CAT ATG ACC AAG GAC CAG
Ser His Ser Lys His His Lys Ala Thr Gly His Met Thr Lys Asp Gln

GTC TTT GTG CCC AAG ACC CGC TCC CAG GTT GGC TTG CCT CCC AAG GAG
Val Phe Val Pro Lys Thr Arg Ser Gln Val Gly Leu Pro Pro Lys Glu
              720
                *
AAC GCT GCT GCT GCC GTT CAG GAG GAG GAC ATG TCC GTG CAC CTG GAT
Asn Ala Ala Ala Ala Val Gln Glu Glu Asp Met Ser Val His Leu Asp
                              780
                                *
GAG GAG GCT CCC ATT GTG ACT TTG TTC TGG ATG GTG ATC CAG TTC TTG
Glu Glu Ala Pro Ile Val Thr Leu Phe Trp Met Val Ile Gln Phe Leu
                                            840
                                              *
TTC GGA TGG CCC GCG TAC CTG ATT ATG AAC GCC TCT GGC CAA GAC TAC
Phe Gly Trp Pro Ala Tyr Leu Ile Met Asn Ala Ser Gly Gln Asp Tyr
```

FIG. 5B

900 *
GGC CGC TGG ACC TCG CAC TTC CAC ACG TAC TCG CCC ATC TTT GAG CCC
Gly Arg Trp Thr Ser His Phe his Thr Tyr Ser Pro Ile Phe Glu Pro

CGC AAC TTT TTC GAC ATT ATT ATC TCG GAC CTC GGT GTG TTG GCT GCC
Arg Asn Phe Phe Asp Ile Ile Ile Ser Asp Leu Gly Val Leu Ala Ala

960 *
CTC GGT GCC CTG ATC TAT GCC TCC ATG CAG TTG TCG CTC TTG ACC GTC
Leu Gly Ala Leu Ile Tyr Ala Ser Met Gln Leu Ser Leu Leu Thr Val

1020 *
ACC AAG TAC TAT ATT GTC CCC TAC CTC TTT GTC AAC TTT TGG TTG GTC
Thr Lys Tyr Tyr Ile Val Pro Tyr Leu Phe Val Asn Phe Trp Lru Val

1080 *
CTG ATC ACC TTC TTG CAG CAC ACC GAT CCC AAG CTG CCC CAT TAC CGC
Leu Ile Thr Phe Leu Gln His Thr Asp Pro Lys Leu Pro His Tyr Arg

1140 *
GAG GGT GCC TGG AAT TTC CAG CGT GGA GCT CTT TGC ACC GTT GAC CGC
Glu Gly Ala Trp Asn Phe Gln Arg Gly Ala Leu Cys Thr Val Asp Arg

TCG TTT GGC AAG TTC TTG GAC CAT ATG TTC CAC GGC ATT GTC CAC ACC
Ser Phe Gly Lys Phe Leu Asp His Met Phe His Gly Ile Val His Thr

1200 *
CAT GTG GCC CAT CAC TTG TTC TCG CAA ATG CCG TTC TAC CAT GCT GAG
His Val Ala His His Leu Phe Ser Gln Met Pro Phe Tyr His Ala Glu

FIG. 5C

```
                                           1260
                                              *
GAA GCT ACC TAT CAT CTC AAG AAA CTG CTG GGA GAG TAC TAT GTG TAC
Glu Ala Thr Tyr His Leu Lys Lys Leu Leu Gly Glu Tyr Tyr Val Tyr
                                                   1320
                                                      *
GAC CCA TCC CCG ATC GTG GTT GCG GTC TGG AGG TCG TTC CGT GAG TGC
Asp Pro Ser Pro Ile Val Val Ala Val Trp Arg Ser Phe Arg Glu Cys
                                                                 1380
                                                                    *
CGA TTC GTG GAG GAT CAG GGA GAC GTG GTC TTT TTC AAG AAG TAAAAA
Arg Phe Val Glu Asp Gln Gly Asp Val Val Phe Phe Lys Lys
                                                                 1440
                                                                    *
AAAAGACAAT GGACCACACA CAACCTTGTC TCTACAGACC TACGTATCAT GTAGCCATAC

CACTTCATAA AAGAACATGA GCTCTAGAGG CGTGTCATTC GCGCCTCC
```

FIG. 5D

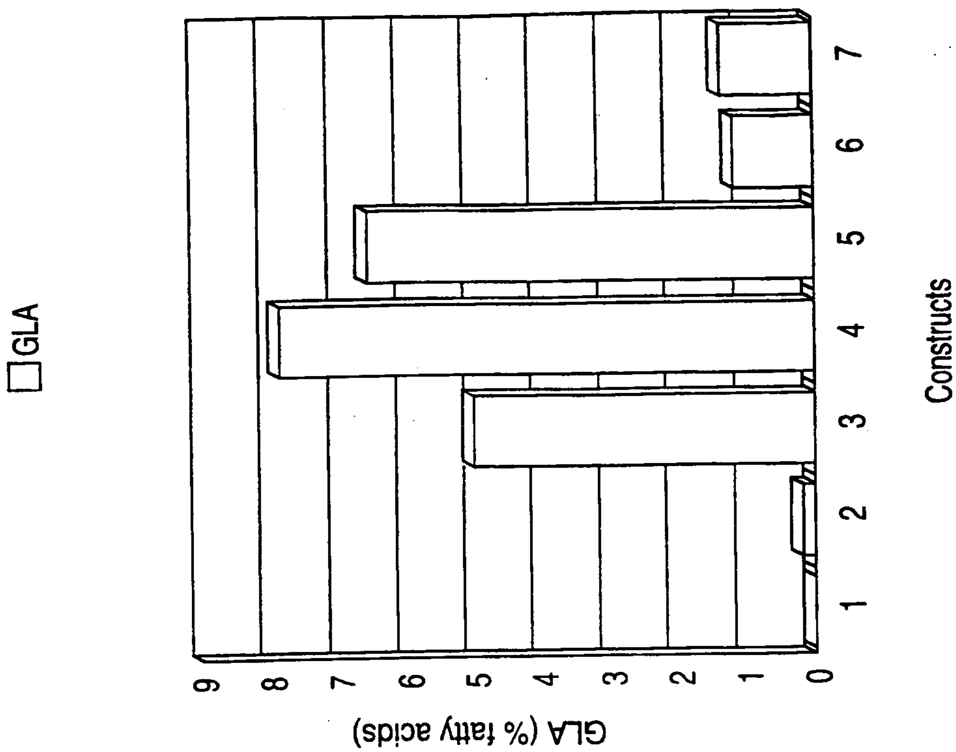

GLA
GLA (% fatty acids)
0
1
2
3
4
5
6
7
8
9
1
2
3
4
5
6
7
Constructs

FIG. 6B

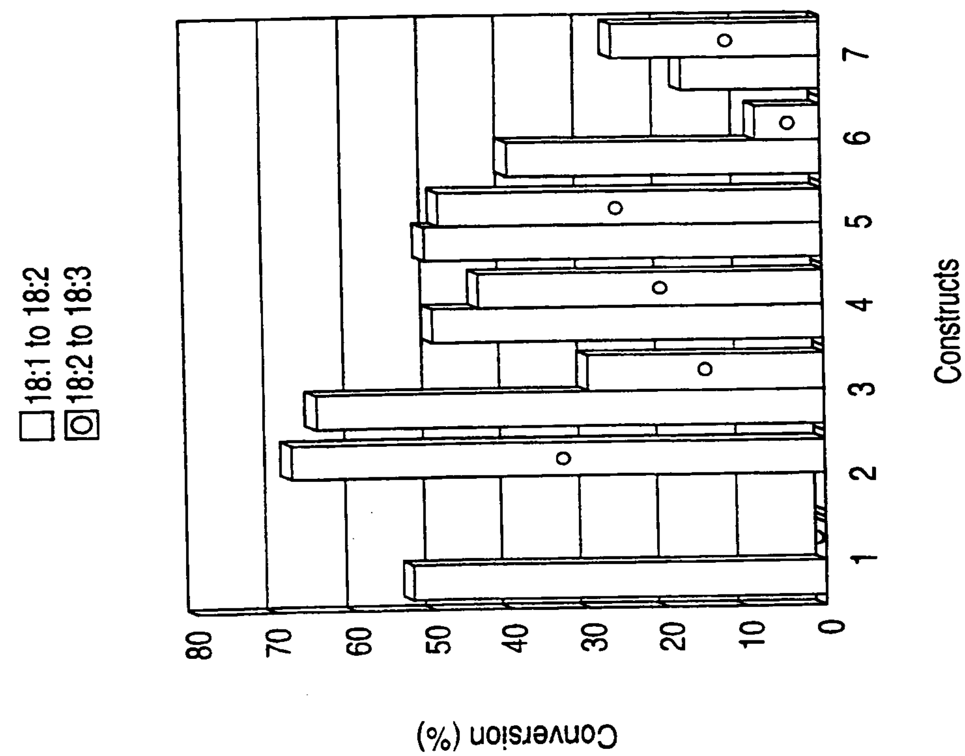

18:1 to 18:2
18:2 to 18:3
Conversion (%)
0
10
20
30
40
50
60
70
80
1
2
3
4
5
6
7
Constructs

FIG. 6A

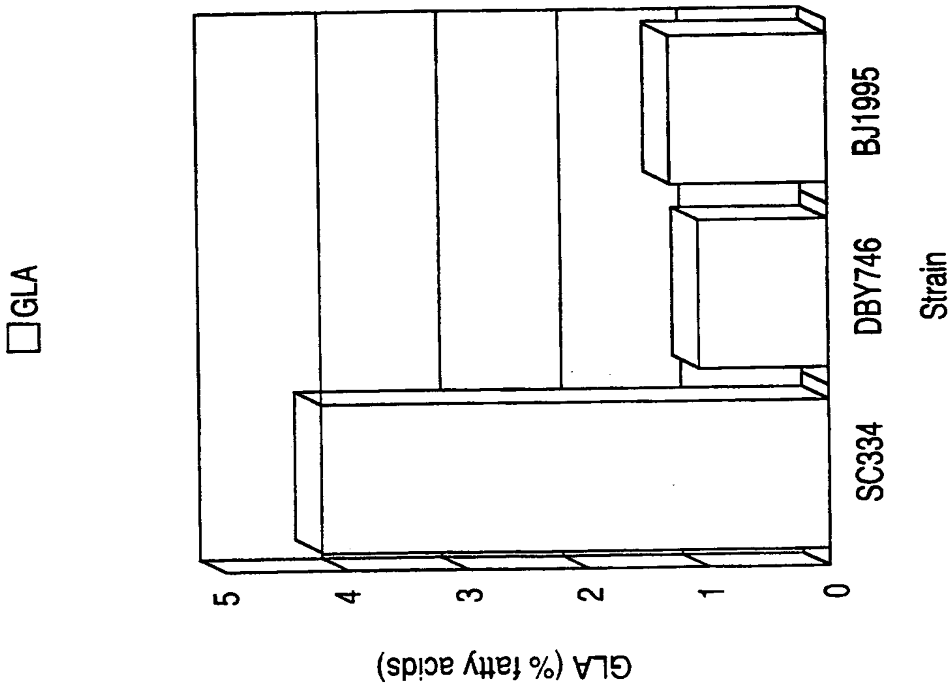

GLA
GLA (% fatty acids)
0
1
2
3
4
5
SC334
DBY746
BJ1995
Strain

FIG. 7B

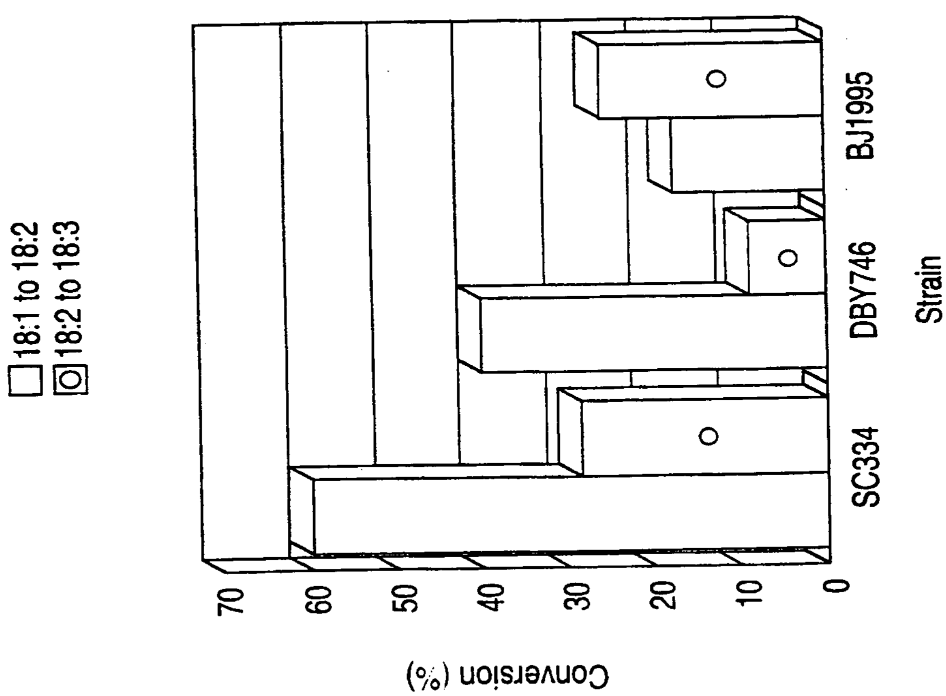

18:1 to 18:2
18:2 to 18:3
Conversion (%)
0
10
20
30
40
50
60
70
SC334
DBY746
BJ1995
Strain

FIG. 7A

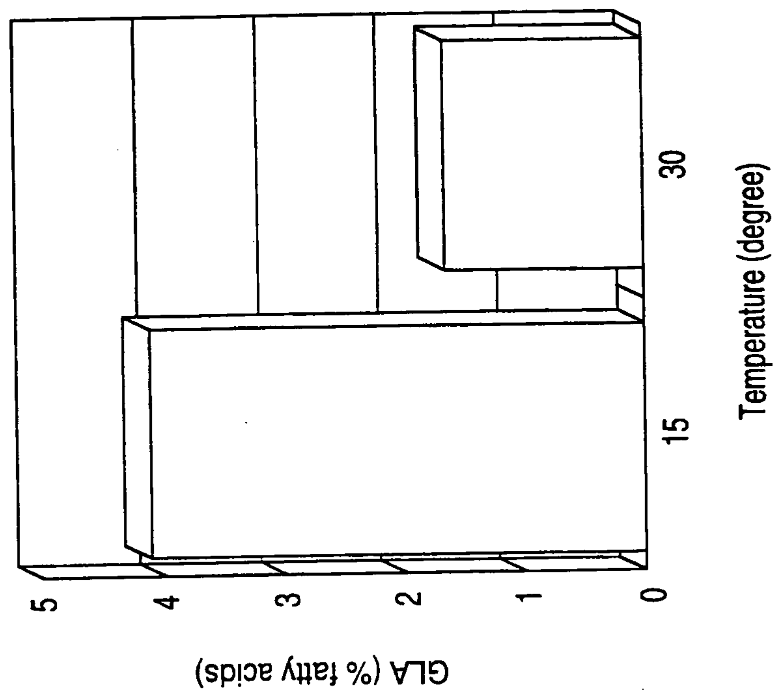

GLA (% fatty acids)
0
1
2
3
4
5
15
30
Temperature (degree)

FIG. 8B

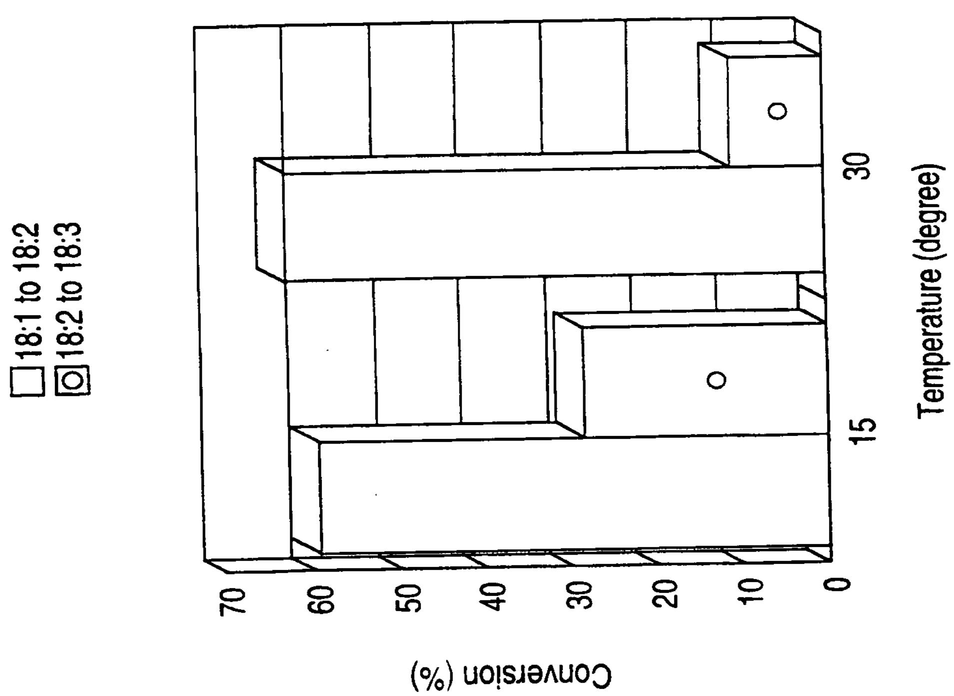

18:1 to 18:2
18:2 to 18:3
Conversion (%)
0
10
20
30
40
50
60
70
15
30
Temperature (degree)

FIG. 8A

```
SCORES   INIT1: 117    INITN: 225    OPT: 256
SMITH-WATERMAN SCORE: 408;  27.0% IDENTITY IN 441 aa OVERLAP

                        10        20        30        40        50
ma29gcg.pep  MGTDQGKT---FTWEELAAHNTKDDLLLAIRGRVYDVTKFLSRHPGGVDTLLLGAGRDVT
                 ||  |   |||:|:|  ::  ::   |:|   :||::::|  |||||  ::   ||:|:|
253538a          QGPTPRYFTWDEVAQRSGCEERWLVIDRKVYNISEFTRRHPGGSRVISHYAGQDAT
                         10        20        30        40        50

             60       70        80        90       100       110
ma29gcg.pep  PVFEMYHAF-GAADAIMKKYYVGTLVSNELPIFPEPTVFHKTIKTRVEGYFTDRNIDPKN
              |  :|   | :   |::  :| | | | | | ||| ::        |   | : :   :
253538a      DPFVAFHINKGLVKKYMNSLLIGEL-SPEQPSF-EPTKNKELTDEFRELRATVERMGLMK
              60       70        80        90       100       110

              120       130       140       150       160       170
ma29gcg.pep  RPEIWGRYALIFGSLIASYYAQLFVPFVVERTWLQVVF-AIIMGFACAQVGLNPLHDASH
              :::  : |  :  :: :: |   :: :|    ::|  :: |:::: : ||:|    || :|
253538a      ANHVF--FLLYLLHILLLDGAAWLTLWVFGTSFLPFLLCAVLLSAVQAQAGWLQ-HDYGH
                120       130       140       150       160       170

              180       190                200       210       220
ma29gcg.pep  FSVTHNPTVWKILGATHDF----FNGASYLVWMYQHMLGHHPYTNIAGADPDVSTSE---
             :|| ::|  |: |  :| |    ::|||  | ::| : ||    ||   ||||:  :
253538a      LSVYRKPK-WNHL--VHKFVIGHLKGASANWWNHRH-FQHHAKPNIFHKDPDVNMLHVFV
                   180       190       200       210       220
```

FIG. 9A

```
SCORES   INIT1: 117    INITN: 225    OPT: 256
SMITH-WATERMAN SCORE: 408;  27.0% IDENTITY IN 441 aa OVERLAP

                   230       240       250         260       270       280
ma29gcg.pep  ----PDVRRIKPNQKWF-VNHINQHMFV--PFLYGLLAFKVRIQDINILYFVKTNDAIRV
                 ::   |  : |::  ||  ::::|:  | |   : |: :|    |: ::  :: : :
253538a      LGEWQPIEYGKKKLKYLPYNHQHEYFFLIGPPLLIPMYFQYQI----IMTMIVHKNWVDL
             230       240       250       260       270           280

                290       300       310       320       330       340
ma29gcg.pep  NPISTWHTVMFWGGKAFFVWYRLIVPLQYLPLGKVLLLFTVADMVSSYWLALTFQANHVV
                :| : ::       ||: |    :|: |   || :||:::    :: |:|:: : | ||:|
253538a      ----AWAVSYYI---RFFITY---IPF-YGILG-ALLFLNFIRFLESHWFVWVTQMNHIV
                 290                300        310       320       330

                350       360            370       380       390
ma29gcg.pep  EEVQWPLPDENGIIQKDWAAMQVETT----QDYAHDSHLWTSITGSLNYQAVHHLFPNVS
              |:      |:::     :||  :  |: :|    |:: :|    |  |   | ||:|  |||||::
253538a      MEI-----DQEAY--RDWFSSQLTATCNVEQSFFND---WFS--GHLNFQIEHHLFPTMP
                       340       350       360            370

              400       410       420       430       440
ma29gcg.pep  QHHYPDILAIIKNTCSEYKVPYLVKDTFWQAFASHLEHLRVLGLRPKEEX
             :|:    |   ::|: |::: :  |  |
253538a      RHNLHKIAPLVKSLCAKHGIEYQEKPLLRALLDIIRSLKKSGKLWLDAYLHKX
            380       390       400       410       420       430
```

FIG. 9B

```
SCORES   INIT1: 231    INITN: 499    OPT: 401
SMITH-WATERMAN SCORE: 620;  27.3% IDENTITY IN 455 aa OVERLAP

                       10        20        30        40        50        59
ma524gcg.pep  MAAAPSVRTFTRAEVLNAEALNEGKKDAEAPFLMIIDNKVYDVREFVPDHPGGSVILTH-
                 |: | ||  ||         :::::     ::|| |||:: ||:  ||||| :::|
253538a          QGPTPRYFTWDEV---------AQRSGCEERWLVIDRKVYNISEFTRRHPGGSRVISHY
                         10                 20        30        40        50

              60        70        80        90            100       110
ma524gcg.pep  VGKDGTDVFDTFHPEAAW--ETLANFYVGDIDE---SDRDIKNDDFAAEVRKLRTLFQSL
              :|:|:|| | :|| : :   : : :: :|:::    | :  || ::: | |:||:  : :
253538a       AGQDATDPFVAFHINKGLVKKYMNSLLIGELSPEQPSFEPTKNKELTDEFRELRATVERM
                     60        70        80        90       100       110

                    120       130       140       150       160       170
ma524gcg.pep  GYYDSSKAYYAFKVSFNLCIWGLSTVIVAKWGQTSTLANVLSAALLGLFWQQCGWLAHDF
              | : :::::: : :   | : | : : :  :| || |   :| |:||:     | ||| ||:
253538a       GLMKANHVFFLLYLLHILLLDGAAWLTLWVFG-TSFLPFLLCAVLLSAVQAQAGWLQHDY
                    120       130       140        150       160

                    180       190       200       210       220       230
ma524gcg.pep  LHHQVFQDRFWGDLFGAFLGGVCQGFSSSWWKDKHNTHHAAPNVHGEDPDIDTHPLLTWS
              | :|::    |: |   |: |  :| |::||: :|  ||| ||:  :|||::    :|
253538a       GHLSVYRKPKWNHLVHKFVIGHLKGASANWWNHRHFQHHAKPNIFHKDPDVN---ML---
              170       180       190       200       210       220
```

FIG. 10A

```
SCORES   INIT1: 231    INITN: 499    OPT: 401
SMITH-WATERMAN SCORE: 620;  27.3% IDENTITY IN 455 aa OVERLAP

                      240       250       260       270          280       290
ma524gcg.pep  EHALEMFSDVPDEELTRMWSRFMVLNQTWFYFPILS---FARLSWCLQSILFVLPNGQAH
               |:: ::::    |  :    :::  |:   || :::    :  :  :   | |: ::    :|
253538a       -HVF-VLGEWQPIEYGKKKLKYLPYNHQHEYFFLIGPPLLIPMYFQYQIIMTMI----VH
                   230       240       250       260       270

                 300       310       320         330       340      349
ma524gcg.pep  KPSGARVPISLVEQLSLAMHWTWYLATMFLFIK--DPVNMLVYFLVSQAVCGNLLAIVFS
              |            : ::||   ::|:  :: :|      :: |:::   : : :: :: | :
253538a       K-----------NWVDLAWAVSYYIRFFITYIPFYGILGALLFLNFIRFLESHWFVWVTQ
                          280       290       300       310       320

             350       360       370       380       390       400      409
ma524gcg.pep  LNHNGMPVISKEEAVDMDFFTKQIITGRDVHPGLFANWFTGGLNYQIEHHLFPSMPRHNF
              :||  | :    :||   |:|::|: :   :|: ::| :||:| ||:||||||||:|||||:
253538a       MNHIVMEI--DQEAYR-DWFSSQLTATCNVEQSFFNDWFSGHLNFQIEHHLFPTMPRHNL
                330       340       350       360       370       380

             410       420       430       440       450
ma524gcg.pep  SKIQPAVETLCKKYNVRYHTTGMIEGTAEVFSRLNEVSKAASKMGKAQX
               || : :::|| |::::|:    ::::  :::  |:: :|
253538a       HKIAPLVKSLCAKHGIEYQEKPLLRALLDIIRSLKKSGKLWLDAYLHKX
                   390       400       410       420       430
```

FIG. 10B

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/07126

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/53 C12N15/81 C12N9/02 C12N5/10 C12N1/19
C12P7/64 C11B1/00 A61K31/20 A23L1/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C12P C11B A61K A23L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	COVELLO P. ET AL.: "Functional expression of the extraplastidial Arabidopsis thaliana oleate desaturase gene (FAD2) in Saccharomyces cerevisiae" PLANT PHYSIOLOGY, vol. 111, no. 1, May 1996, pages 223-226, XP002075211 see the whole document ---	10
X	WO 94 11516 A (DU PONT ;LIGHTNER JONATHAN EDWARD (US); OKULEY JOHN JOSEPH (US)) 26 May 1994 cited in the application	10
A	see the whole document ---	1-9, 11-98

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 August 1998	03/09/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Kania, T

1

Int tional Application No
PCT/US 98/07126

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	WO 93 06712 A (RHONE POULENC AGROCHIMIE) 15 April 1993 cited in the application see the whole document	10,65-67
X	WO 96 21022 A (RHONE POULENC AGROCHIMIE) 11 July 1996 cited in the application * see the whole document, esp. p. 2 1.3-21 *	10,65-92
X	WO 94 18337 A (MONSANTO CO ;UNIV MICHIGAN (US); GIBSON SUSAN IRMA (US); KISHORE G) 18 August 1994 * see the whole document, esp. claims 8-10 *	10, 57-59, 65-92, 97,98
X	EP 0 561 569 A (LUBRIZOL CORP) 22 September 1993 cited in the application see the whole document	57-59, 65-92, 97,98
P,X	WO 97 30582 A (CARNEGIE INST OF WASHINGTON ;MONSANTO COMPANY INC (US); BROUN PIER) 28 August 1997 see the whole document	10
P,X	YOSHINO R. ET AL.: "Developmental cDNA in Dictyostelium discoideum, AC C25549" EMBL DATABASE, 24 July 1997, XP002075237 Heidelberg see the whole document	96

1

International application No.

PCT/US 98/07126

INTERNATIONAL SEARCH REPORT

Box I Observations wher certain claims w re found unsearchable (Continuati n of it m 1 f first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claims 68, 87, 88 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [X] Claims Nos.: (not applicable)
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [X] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This international Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-94, 97, 98

Isolated nucleic acids comprising SEQ ID NO: 1,3, as well as polypeptides comprising SEQ ID NO: 2,4, homologs and fragments thereof.
An isolated or purified eukaryotic polypeptide which desaturates a fatty acid molecule at carbon 6 or 12, especially of fungal origin, especially of Mortierella alpina.
Nucleic acid constructs and vectors comprising delta-6, or delta 12 desaturases according to SEQ ID NO: 1,3, derived from the fungus Mortierella alpina.
Recombinant cells comprising said constructs.
Methods for the production of GLA, stearidonic acid, linoleic acid, or gamma-linolenic acid in eukaryotic cell cultures, especially yeast cultures, employing DNA sequences or constructs coding for delta-6, or delta-12 desaturases of fungal origin, especially of Mortierella alpina.
Methods for obtaining altered long chain polyunsaturated fatty acid biosynthesis using plants comprising delta-6, or delta-12 desaturases, or combinations thereof, derived from fungi or algae.
Plant oils derived from said plants and their use for therapeutical, nutritional, and cosmetical purposes, as well as products derived therefrom.

2. Claim : 95

An isolated peptides sequence selected from the group of SEQ ID NO: 34-40.

3. Claim : 96

An isolated peptides sequence selected from the group consisting of SEQ ID NO: 20, 22, 25, 26

Claims No.: not applicable

In view of the extremely broad claims 5-8, the search was executed with due regard to the PCT Search guidelines (PCT/GL/2), C-III, paragraph 2.2, 2.3 read in conjuction with 3.7 and Rule 33.3 PCT, i.e. particular emphasis was put on the inventive concept, as illustrated by Mortierella alpina fatty acid desaturases comprising the nucleotide sequences in SEQ ID NO:1 and 3.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 98/07126

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
WO 9411516	A	26-05-1994	AU	5407594	A	08-06-1994
			CA	2149223	A	26-05-1994
			EP	0668919	A	30-08-1995
			JP	8503364	T	16-04-1996
WO 9306712	A	15-04-1993	AU	667848	B	18-04-1996
			AU	2881292	A	03-05-1993
			BG	98695	A	31-05-1995
			BR	9206613	A	11-04-1995
			CA	2120629	A	15-04-1993
			CN	1072722	A	02-06-1993
			CN	1174236	A	25-02-1998
			CZ	9400817	A	13-09-1995
			EP	0666918	A	16-08-1995
			HU	69781	A	28-09-1995
			JP	7503605	T	20-04-1995
			MX	9205820	A	01-04-1993
			NZ	244685	A	27-06-1994
			US	5552306	A	03-09-1996
			US	5614393	A	25-03-1997
			US	5689050	A	18-11-1997
			US	5663068	A	02-09-1997
			US	5789220	A	04-08-1998
			ZA	9207777	A	21-04-1993
WO 9621022	A	11-07-1996	US	5614393	A	25-03-1997
			AU	4673596	A	24-07-1996
			CA	2207906	A	11-07-1996
			CN	1177379	A	25-03-1998
			EP	0801680	A	22-10-1997
			US	5789220	A	04-08-1998
WO 9418337	A	18-08-1994	EP	0684998	A	06-12-1995
			JP	8506490	T	16-07-1996
EP 0561569	A	22-09-1993	AU	3516793	A	16-09-1993
			CA	2092661	A	14-09-1993
			JP	6014667	A	25-01-1994
			US	5777201	A	07-07-1998

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int tional Application No
PCT/US 98/07126

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9730582 A	28-08-1997	AU 2050497 A	10-09-1997

Form PCT/ISA/210 (patent family annex) (July 1992)

[redacted]
28

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/53, 15/83, 5/10, C12P 7/64, C11B 1/00, A61K 31/20, A23L 1/30, A23K 1/00

A1

(11) International Publication Number: WO 98/46765

(43) International Publication Date: 22 October 1998 (22.10.98)

(21) International Application Number: PCT/US98/07422

(22) International Filing Date: 10 April 1998 (10.04.98)

(30) Priority Data:
08/833,610 11 April 1997 (11.04.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 08/833,610 (CIP)
Filed on 11 April 1997 (11.04.97)

(71) Applicants (*for all designated States except US*): CALGENE LLC [US/US]; 1920 Fifth Street, Davis, CA 95616 (US). ABOTT LABORATORIES [US/US]; 100 Abbott Park Road, Abbott Park, IL 60064-3500 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): KNUTZON, Deborah [US/US]; 6110 Rockhurst Way, Granite Bay, CA 95746 (US). MUKERJI, Pradip [US/US]; 1069 Arcaro Drive, Gahanna, OH 43230 (US). HUANG, Yung-Sheng [CA/US]; 2462 Danvers Court, Upper Arlington, OH 43220 (US). THURMOND, Jennifer [US/US]; 3702 Adirondack, Columbus, OH 43231 (US). CHAUDHARY, Sunita [IN/US]; 3419 Woodbine Place, Pearland, TX 77584 (US). LEONARD, Amanda, Eun-Yeong [US/US]; 581 Shadewood Court, Gahanna, OH 43230 (US).

(74) Agents: WARD, Michael, R. et al.; Limbach & Limbach L.L.P., 2001 Ferry Building, San Francisco, CA 94111-4262 (US).

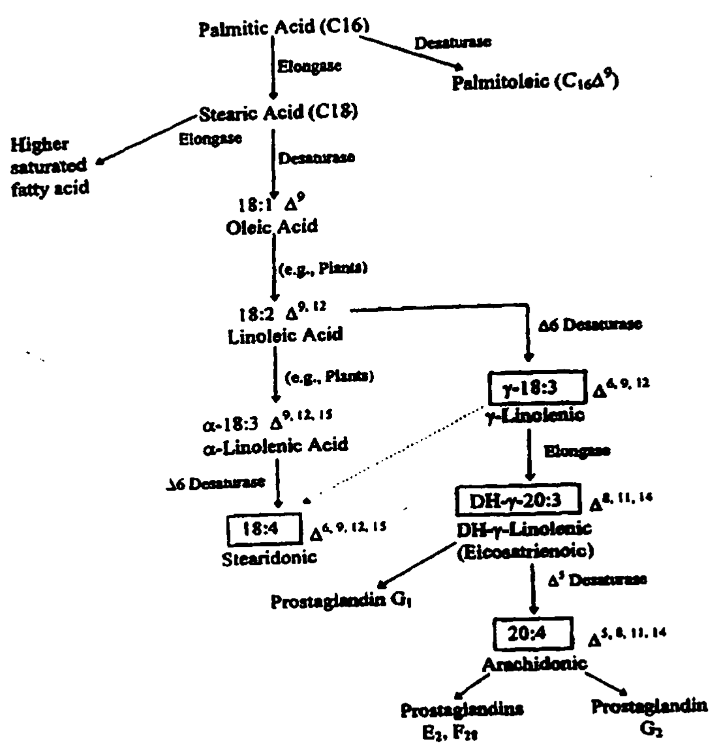
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHODS AND COMPOSITIONS FOR SYNTHESIS OF LONG CHAIN POLYUNSATURATED FATTY ACIDS

(57) Abstract

The present invention relates to a fatty acid Δ5-desaturase able to catalyze the conversion of dihomo-gamma-linolenic acid to arachidonic acid. Nucleic acid sequences encoding Δ5-desaturase, nucleic acid sequences which hybridize thereto, DNA constructs comprising a Δ5-desaturase gene, and recombinant host microorganism or animal expressing increased levels of a Δ5-desaturase are described. Methods for desaturating a fatty acid at the Δ5 position and for producing arachidonic acid by expressing increased levels of a Δ5-desaturase are disclosed. Fatty acids, and oils containing them, which have been desaturated by a Δ5-desaturase produced by recombinant host microorganisms or animals are provided. Pharmaceutical compositions, infant formulas or dietary supplements containing fatty acids which have been desaturated by a Δ5-desaturase produced by a recombinant host microorganism or animal also are described.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

METHODS AND COMPOSITIONS FOR SYNTHESIS OF LONG CHAIN POLYUNSATURATED FATTY ACIDS

RELATED APPLICATION

This application is a continuation in part application of Serial Number 08/833,610 filed April 11, 1997.

INTRODUCTION

Field of the Invention

This invention relates to modulating levels of enzymes and/or enzyme components relating to production of long chain poly-unsaturated fatty acids (PUFAs) in a microorganism or animal.

Background

Two main families of polyunsaturated fatty acids (PUFAs) are the ω3 fatty acids, exemplified by eicosapentaenoic acid (EPA), and the ω6 fatty acids, exemplified by arachidonic acid (ARA). PUFAs are important components of the plasma membrane of the cell, where they may be found in such forms as phospholipids. PUFAs are necessary for proper development, particularly in the developing infant brain, and for tissue formation and repair. PUFAs also serve as precursors to other molecules of importance in human beings and animals, including the prostacyclins, eicosanoids, leukotrienes and prostaglandins.

Four major long chain PUFAs of importance include docosahexaenoic acid (DHA) and EPA, which are primarily found in different types of fish oil, gamma-linolenic acid (GLA), which is found in the seeds of a number of plants, including evening primrose (*Oenothera biennis*), borage (*Borago officinalis*) and black currants (*Ribes nigrum*), and stearidonic acid (SDA), which is found in marine oils and plant seeds. Both GLA and another important long chain PUFA, arachidonic acid (ARA), are found in filamentous fungi. ARA can be

purified from animal tissues including liver and adrenal gland. GLA, ARA, EPA and SDA are themselves, or are dietary precursors to, important long chain fatty acids involved in prostaglandin synthesis, in treatment of heart disease, and in development of brain tissue.

Polyunsaturated fatty acids have a number of pharmaceutical and medical applications including treatment of heart disease, cancer and arthritis.

For DHA, a number of sources exist for commercial production including a variety of marine organisms, oils obtained from cold water marine fish, and egg yolk fractions. For ARA, microorganisms including the genera *Mortierella, Entomophthora, Phytium* and *Porphyridium* can be used for commercial production. Commercial sources of SDA include the genera *Trichodesma and Echium*. Commercial sources of GLA include evening primrose, black currants and borage. However, there are several disadvantages associated with commercial production of PUFAs from natural sources. Natural sources of PUFAs, such as animals and plants, tend to have highly heterogeneous oil compositions. The oils obtained from these sources therefore can require extensive purification to separate out one or more desired PUFAs or to produce an oil which is enriched in one or more PUFA. Natural sources also are subject to uncontrollable fluctuations in availability. Fish stocks may undergo natural variation or may be depleted by overfishing. Fish oils have unpleasant tastes and odors, which may be impossible to economically separate from the desired product, and can render such products unacceptable as food supplements. Animal oils, and particularly fish oils, can accumulate environmental pollutants. Weather and disease can cause fluctuation in yields from both fish and plant sources. Cropland available for production of alternate oil-producing crops is subject to competition from the steady expansion of human populations and the associated increased need for food production on the remaining arable land. Crops which do produce PUFAs, such as borage, have not been adapted to commercial growth and may not perform well in monoculture. Growth of such crops is thus not economically competitive where more profitable and better established crops can be grown. Large scale

fermentation of organisms such as *Mortierella* is also expensive. Natural animal tissues contain low amounts of ARA and are difficult to process. Microorganisms such as *Porphyridium* and *Mortierella* are difficult to cultivate on a commercial scale.

Dietary supplements and pharmaceutical formulations containing PUFAs can retain the disadvantages of the PUFA source. Supplements such as fish oil capsules can contain low levels of the particular desired component and thus require large dosages. High dosages result in ingestion of high levels of undesired components, including contaminants. Unpleasant tastes and odors of the supplements can make such regimens undesirable, and may inhibit compliance by the patient. Care must be taken in providing fatty acid supplements, as overaddition may result in suppression of endogenous biosynthetic pathways and lead to competition with other necessary fatty acids in various lipid fractions *in vivo*, leading to undesirable results. For example, Eskimos having a diet high in ω3 fatty acids have an increased tendency to bleed (U.S. Pat. No. 4,874,603).

A number of enzymes are involved in PUFA biosynthesis. Linolenic acid (LA, 18:2 Δ9, 12) is produced from oleic acid (18:1 Δ^{9}) by a Δ12-desaturase. GLA (18:3 Δ6, 9, 12) is produced from linoleic acid (LA, 18:2 Δ9, 12) by a Δ6-desaturase. ARA (20:4 Δ5, 8, 11, 14) production from dihomo-gamma-linolenic acid (DGLA, 20:3 Δ8, 11, 14) is catalyzed by a Δ5-desaturase. However, animals cannot desaturate beyond the Δ9 position and therefore cannot convert oleic acid (18:1 Δ9) into linolenic acid (18:2 Δ912). Likewise, α-linoleic acid (ALA, 18:3 Δ9, 12, 15) cannot be synthesized by mammals. Other eukaryotes, including fungi and plants, have enzymes which desaturate at positions Δ12 and Δ15. The major poly-unsaturated fatty acids of animals therefore are either derived from diet and/or from desaturation and elongation of linoleic acid (18:2 Δ9, 12) or ∝-linolenic acid (18:3 Δ9, 12, 15). Therefore it is of interest to obtain genetic material involved in PUFA biosynthesis from species that naturally produce these fatty acids and to express the isolated

material in a microbial or animal system which can be manipulated to provide production of commercial quantities of one or more PUFAs. Thus there is a need for fatty acid desaturases, genes encoding them, and recombinant methods of producing them. A need further exists for oils containing higher relative proportions of and/or enriched in specific PUFAs. A need also exists for reliable economical methods of producing specific PUFAs.

Relevant Literature

Production of gamma-linolenic acid by a Δ6-desaturase is described in USPN 5,552,306. Production of 8, 11-eicosadienoic acid using *Mortierella alpina* is disclosed in USPN 5,376,541. Production of docosahexaenoic acid by dinoflagellates is described in USPN 5,407,957. Cloning of a Δ6-palmitoyl-acyl carrier protein desaturase is described in PCT publication WO 96/13591 and USPN 5,614,400. Cloning of a Δ6-desaturase from borage is described in PCT publication WO 96/21022. Cloning of Δ9-desaturases is described in the published patent applications PCT WO 91/13972, EP 0 550 162 A1, EP 0 561 569 A2, EP 0 644 263 A2, and EP 0 736 598 A1, and in USPN 5,057,419. Cloning of Δ12-desaturases from various organisms is described in PCT publication WO 94/11516 and USPN 5,443,974. Cloning of Δ15-desaturases from various organisms is described in PCT publication WO 93/11245. All publications and U.S. patents or applications referred to herein are hereby incorporated in their entirety by reference.

Summary of the Invention

Novel compositions and methods are provided for preparation of poly-unsaturated long chain fatty acids or PUFAs. The compositions include nucleic acids encoding a Δ5-desaturase and/or polypeptides having Δ5-desaturase activity, the polypeptides, and probes for isolating and detecting the same. The methods involve growing a host microorganism or animal which contains and expresses one or more transgenes encoding a Δ5-desaturase and/or a polypeptide having Δ5-desaturase activity. Expression of the desaturase

polypeptide provides for a relative increase in Δ5-desaturated PUFA, or metabolic progeny therefrom, as a result of altered concentrations of enzymes and substrates involved in PUFA biosynthesis. The invention finds use for example in the large scale production of PUFA containing oils which include, for example, ARA, EPA and/or DHA.

In a preferred embodiment, a nucleic acid sequence comprising a Δ5-desaturase depicted in Figure 3A-D (SEQ ID NO 1), a polypeptide encoded by the nucleic acid, and a purified or isolated polypeptide depicted in Figure 3A-D (SEQ ID NO: 2), and an isolated nucleic acid encoding the polypeptide of Figure 3A-D (SEQ ID NO: 2) are provided. Another embodiment of the invention is an isolated nucleic acid sequence which encodes a polypeptide, wherein said polypeptide desaturates a fatty acid molecule at carbon 5 from the carboxyl end of the molecule. The nucleic acid is preferably derived from a eukaryotic cell, such as a fungal cell, or a fungal cell of the genus *Mortierella*, or of the genus/species *Mortierella alpina*. Also preferred is an isolated nucleic acid comprising a sequence which anneals to a nucleotide sequence depicted in Figure 3A-3D (SEQ ID NO: 1), and a nucleic acid which encodes an amino acid sequence depicted in Figure 3A-D (SEQ ID NO: 2). In particular, the nucleic acid encodes an amino acid sequence depicted in Figure 3A-D (SEQ ID NO: 2) which is selected from the group consisting of amino acid residues 30-38, 41-44, 171-175, 203-212, and 387-394. In an additional embodiment, the invention provides an isolated or purified polypeptide which desaturates a fatty acid molecule at carbon 5 from the carboxyl end of the molecule. Also provided is an isolated nucleic acid sequence which hybridizes to a nucleotide sequence depicted in Figure 3A-D (SEQ ID NO 1), an isolated nucleic acid sequence having at least about 50% identity to Figure 3A-D (SEQ ID NO 1).

The present invention further includes a nucleic acid construct comprising a nucleotide sequence depicted in a Figure 3A-D (SEQ ID NO: 1) linked to a heterologous nucleic acid; a nucleic acid construct comprising a nucleotide sequence depicted in a Figure 3A-D (SEQ ID NO: 1) operably linked to a promoter; and a nucleic acid construct comprising a nucleotide sequence

depicted in a Figure 3A-D (SEQ ID NO: 1) operably linked to a promoter which is functional in a microbial cell. In a preferred embodiment, the microbial cell is a yeast cell, and the nucleotide sequence is derived from a fungus, such as a fungus of the genus *Mortierella*, particularly a fungus of the species *Mortierella alpina*.

In another embodiment of the invention, a nucleic acid construct is provided which comprises a nucleotide sequence which encodes a polypeptide comprising an amino acid sequence which corresponds to or is complementary to an amino acid sequence depicted in Figure 3A-D (SEQ ID NO: 2), wherein the nucleotide sequence is operably linked to a promoter which is functional in a host cell, and wherein the nucleotide sequence encodes a polypeptide which desaturates a fatty acid molecule at carbon 5 from the carboxyl end of a fatty acid molecule. Additionally, provided by the invention is a nucleic acid construct comprising a nucleotide sequence which encodes a functionally active Δ5-desaturase, where the desaturase includes an amino acid sequence which corresponds to or is complementary to all of or a portion of an amino acid sequence depicted in a Figure 3A-D (SEQ ID NO: 2), wherein the nucleotide sequence is operably linked to a promoter functional in a host cell.

The invention also includes a host cell comprising a nucleic acid construct of the invention. In a preferred embodiment, a recombinant host cell is provided which comprises at least one copy of a DNA sequence which encodes a functionally active *Mortierella alpina* fatty acid desaturase having an amino acid sequence as depicted in Figure 3A-D (SEQ ID NO: 2), wherein the cell or an ancestor of the cell was transformed with a vector comprising said DNA sequence, and wherein the DNA sequence is operably linked to a promoter. The host cell is either eukaryotic or prokaryotic. Preferred eukaryotic host cells are those selected from the group consisting of a mammalian cell, an insect cell, a fungal cell, and an algae cell. Preferred mammalian cells include an avian cell, a fungal cell such as a yeast, and a marine algae cell. Preferred prokaryotic cells include those selected from the group consisting of a bacteria, a cyanobacteria, cells which contain a

bacteriophage, and/or a virus. The DNA sequence of the recombinant host cell preferably contains a promoter which is functional in the host cell.

The host cells of the invention which contain the DNA sequences of the invention are enriched for fatty acids, such as 20:3 fatty acids. In a preferred embodiment, the host cells are enriched for 20:4 fatty acids as compared to an untransformed host cell which is devoid of said DNA sequence, and/or enriched for 20:5 fatty acids compared to an untransformed host cell which is devoid of said DNA sequence. In yet another preferred embodiment, the invention provides a recombinant host cell which comprises a fatty acid selected from the group consisting of a dihomo-γ-linolenic acid, n-6 eicosatrienoic acid, 20:3n-6 acid and 20:3 (8,11,14) acid.

The present invention also includes method for production of arachidonic acid in a microbial cell culture, where the method comprises growing a microbial cell culture having a plurality of microbial cells which contain one or more nucleic acids encoding a polypeptide which converts dihomo-γ-linolenic acid to arachidonic acid, wherein the nucleic acid is operably linked to a promoter, under conditions whereby said one or more nucleic acids are expressed, whereby arachidonic acid is produced in the microbial cell culture. In several preferred embodiments of the invention, the polypeptide is an enzyme which desaturates a fatty acid molecule at carbon 5 from the carboxyl end of the fatty acid molecule; the nucleic acid is derived from a *Mortierella sp.*; and the substrate for said polypeptide is exogenously supplied. The microbial cells used in the methods can be either eukaryotic cells or prokaryotic cells. The preferred eukaryotic cells are those selected from the group consisting of a mammalian cell, an insect cell, a fungal cell, and an algae cell. Preferred mammalian cells include an avian cell, a preferred fungal cell is a yeast, and the preferred algae cell is a marine algae cell. The preferred prokaryotic cells include those selected from the group consisting of a bacteria, a cyanobacteria, cells which contain a bacteriophage, and/or a virus. The nucleic acid sequence encoding the polypeptide of the microbial cell preferably contains a promoter which is functional in the host cell which

optionally is an inducible promoter for example by components of the culture broth. The preferred microbial cells used in the methods are yeast cells, such as *Saccharomyces* cells.

In another embodiment of the invention, a recombinant yeast cell is provided which converts greater than about 5% of 20:3 fatty acid substrate to a 20:4 fatty acid product.

Also provided is an oil comprising one or more PUFA. The amount of said one or more PUFAs is approximately 0.3-30% arachidonic acid (ARA), approximately 0.2-30% dihomo-γ-linolenic acid (DGLA), and approximately 0.2-30% γ-linolenic acid (GLA). A preferred oil of the invention is one in which the ratio of ARA:DGLA:GLA is approximately 1.0:19.0:30 to 6.0:1.0:0.2. Another preferred embodiment of the invention is a pharmaceutical composition comprising the oils in a pharmaceutically acceptable carrier. Further provided is a nutritional composition comprising the oils of the invention. The nutritional compositions of the invention preferably are administered to a mammalian host parenterally or internally. A preferred composition of the invention for internal consumption is an infant formula. In a preferred embodiment, the nutritional compositions of the invention are in a liquid form or a solid form.

The present invention also includes a method for desaturating a fatty acid, where the method comprises culturing a recombinant microbial cell of the invention under conditions suitable for expression of a polypeptide encoded by the nucleic acid, wherein the host cell further comprises a fatty acid substrate of the polypeptide. In a preferred embodiment, a fatty acid desaturated by the methods is provided, including an oil comprising the fatty acid.

The present invention is also directed to purified nucleotide and peptide sequences presented in SEQ ID NO:1-34. The present invention is further directed toward methods of using the sequences presented in SEQ ID NO:1-34 as probes to identify related sequences, as components of expression systems and as components of systems useful for producing transgenic oil.

The present invention is further directed to methods of obtaining altered long chain poly unsaturated fatty acid biosystems by growing transgenic microbes which encode transgene expression products which desaturate a fatty acid molecule at carbon 5 from the carboxyl end of the fatty acid molecule.

The present invention is further directed to formulas, dietary supplements or dietary supplements in the form of a liquid or a solid containing the long chain fatty acids of the invention. These formulas and supplements may be administered to a human or an animal.

The formulas and supplements of the invention may further comprise at least one macronutrient selected from the group consisting of coconut oil, soy oil, canola oil, mono- and diglycerides, glucose, edible lactose, electrodialysed whey, electrodialysed skim milk, milk whey, soy protein, and other protein hydrolysates.

The formulas of the present invention may further include at least one vitamin selected from the group consisting of Vitamins A, C, D, E, and B complex; and at least one mineral selected from the group consisting of calcium, magnesium, zinc, manganese, sodium, potassium, phosphorus, copper, chloride, iodine, selenium, and iron.

The present invention is further directed to a method of treating a patient having a condition caused by insuffient intake or production of polyunsaturated fatty acids comprising administering to the patient a dietary substitute of the invention in an amount sufficient to effect treatment of the patient.

The present invention is further directed to cosmetic and pharmaceutical compositions of the material of the invention.

The present invention is also directed to an isolated nucleotide sequence comprising a nucleuotide sequence selected from the group consisting of: SEQ ID NO:13; SEQ ID NO:15; SEQ ID NO:17; SEQ ID NO:19; SEQ ID NO:21; SEQ ID NO:22; SEQ ID NO:22; SEQ ID NO:24; SEQ ID NO:25; SEQ ID NO:26 and SEQ ID NO:27.

The present invention is also directed to an isolated peptide sequence comprising a peptide sequence selected from the group consisting of: SEQ ID NO:14; SEQ ID NO:16; SEQ ID NO:18; SEQ ID NO:20; SEQ ID NO:28; SEQ ID NO:29; SEQ ID NO:30; SEQ ID NO:31; SEQ ID NO:32; SEQ ID NO:33 and SEQ ID NO:34.

The present invention is further directed to transgenic oils in pharmaceutically acceptable carriers. The present invention is further directed to nutritional supplements, cosmetic agents and infant formulae containing transgenic oils.

The present invention is further directed to a method for obtaining altered long chain polyunsaturated fatty acid biosynthesis comprising the steps of: growing a microbe having cells which contain a transgene which encodes a transgene expression product which desaturates a fatty acid molecule at carbon 5 from the carboxyl end of said fatty acid molecule, wherein the trangene is operably associated with an expression control sequence, under conditions whereby the transgene is expressed, whereby long chain polyunsaturated fatty acid biosynthesis in the cells is altered.

The present invention is further directed to the use of chain polyunsaturated fatty acid selected from the group consisting of ARA, DGLA and EPA.

The present invention is further directed toward pharmaceutical compositions comprising at least one nutrient selected from the group consisting of a vitamin, a mineral, a carbohydrate, a sugar, an amino acid, a free fatty acid, a phospholipid, an antioxidant, and a phenolic compound.

Brief Description of the Drawings

Figure 1 shows possible pathways for the synthesis of arachidonic acid (20:4 Δ5, 8, 11, 14) and stearidonic acid (18:4 Δ6, 9, 12, 15) from palmitic acid (C_{16}) from a variety of organisms, including algae, *Mortierella* and humans. These PUFAs can serve as precursors to other molecules important for humans and other animals, including prostacyclins, leukotrienes, and prostaglandins, some of which are shown.

Figure 2 shows possible pathways for production of PUFAs in addition to ARA, including EPA and DHA, for a variety of organisms.

Figure 3A-D shows the DNA sequence of the *Mortierella alpina* Δ5-desaturase and the deduced amino acid sequence.

Figure 4 shows the deduced amino acid sequence of the PCR fragment (see Example 1)

Figure 5A and 5B show alignments of the protein sequence of the Δ5-desaturase with Δ6-desaturases.

Figure 6A and 6B show the effect of the timing of substrate addition relative to induction on conversion of substrate to product in SC334 containing the Δ5-desaturase gene.

Figure 7A and 7B show the effect of inducer concentration on Δ5-desaturase expression in SC334.

Figure 8A and 8B show the effect of induction temperature on Δ5-desaturase activity in SC334.

Figure 9A and 9B show the effect of host strain on the conversion of substrate to product in strains expressing the Δ5-desaturase gene at 15°C.

Figure 10A and 10B show the effect of host strain on the conversion of substrate to product in strains expressing the Δ5-desaturase gene at 30°C.

Figure 11 shows the effect of a host strain expressing choline transferase as well as the Δ5-desaturase gene on the conversion of substrate to product.

Figure 12A and 12B show the effect of media composition and temperature on the conversion of substrate to product in two host strains expressing the Δ5-desaturase gene.

Figure 13 shows alignment of the protein sequence of Ma 29 and contig 253538a.

Figure 14 shows alignment of the protein sequence of Ma 524 and contig 253538a.

Brief Description of the Sequence Listings

SEQ ID NO:1 shows a DNA sequence of the *Mortierella alpina* Δ5-desaturase.

SEQ ID NO:2 shows an amino acid sequence of *Mortierella alpina* Δ5-desaturase.

SEQ ID NO: 3 shows the deduced amino acid sequence of the *M. alpina* PCR fragment (see Example 1).

SEQ ID NO: 4 - SEQ ID NO: 7 show the deduced amino acid sequences of various Δ6-desaturases.

SEQ ID NO: 8 and SEQ ID NO: 9 show PCR primer sequences for Δ6-desaturases

SEQ ID NO: 10 shows a primer for reverse transcription of total RNA.

SEQ ID NO: 11 and SEQ ID NO: 12 show amino acid motifs for desaturase sequences.

SEQ ID NO: 13 and SEQ ID NO: 14 show the nucleotide and amino acid sequence of a *Dictyostelium discoideum* desaturase sequence.

SEQ ID NO: 15 and SEQ ID NO: 16 show the nucleotide and amino acid sequence of a *Phaeodactylum tricornutum* desaturase sequence.

SEQ ID NO: 17-20 show the nucleotide and deduced amino acid sequence of a Schizochytrium cDNA clone.

SEQ ID NO: 21-27 show nucleotide sequences for human desaturases.

SEQ ID NO: 28 - SEQ ID NO: 34 show peptide sequences for human desaturases.

Detailed Description of the Invention

In order to ensure a complete understanding of the invention, the following definitions are provided:

Δ5-Desaturase: Δ5 desaturase is an enzyme which introduces a double bond between carbons 5 and 6 from the carboxyl end of a fatty acid molecule.

Δ6-Desaturase: Δ6-desaturase is an enzyme which introduces a double bond between carbons 6 and 7 from the carboxyl end of a fatty acid molecule.

Δ9-Desaturase: Δ9-desaturase is an enzyme which introduces a double bond between carbons 9 and 10 from the carboxyl end of a fatty acid molecule.

Δ12-Desaturase: Δ12-desaturase is an enzyme which introduces a double bond between carbons 12 and 13 from the carboxyl end of a fatty acid molecule.

Fatty Acids: Fatty acids are a class of compounds containing a long hydrocarbon chain and a terminal carboxylate group. Fatty acids include the following:

Fatty Acid		
12:0	lauric acid	
16:0	palmitic acid	
16:1	palmitoleic acid	

Fatty Acid		
18:0	stearic acid	
18:1	oleic acid	Δ9-18:1
18:2 Δ5,9	taxoleic acid	Δ5,9-18:2
18:2 Δ6,9	6,9-octadecadienoic acid	Δ6,9-18:2
18:2	linoleic acid	Δ9,12-18:2 (LA)
18:3 Δ6,9,12	gamma-linolenic acid	Δ6,9,12-18:3 (GLA)
18:3 Δ5,9,12	pinolenic acid	Δ5,9,12-18:3
18:3	alpha-linolenic acid	Δ9,12,15-18:3 (ALA)
18:4	stearidonic acid	Δ6,9,12,15-18:4 (SDA)
20:0	Arachidic acid	
20:1	Eicoscenic Acid	
22:0	behehic acid	
22:1	erucic acid	
22:2	Docasadienoic acid	
20:4 ω6	arachidonic acid	Δ5,8,11,14-20:4 (ARA)
20:3 ω6	ω6-eicosatrienoic dihomo-gamma linolenic	Δ8,11,14-20:3 (DGLA)
20:5 ω3	Eicosapentanoic (Timnodonic acid)	Δ5,8,11,14,17-20:5 (EPA)
20:3 ω3	ω3-eicosatrienoic	Δ11,16,17-20:3
20:4 ω3	ω3-eicosatetraenoic	Δ8,11,14,17-20:4
22:5 ω3	Docasapentaenoic	Δ7,10,13,16,19-22:5 (ω3DPA)
22:6 ω3	Docosahexaenoic (cervonic acid)	Δ4,7,10,13,16,19-22:6 (DHA)
24:0	Lignoceric acid	

Taking into account these definitions, the present invention is directed to novel DNA sequences, DNA constructs, methods and compositions are provided which permit modification of the poly-unsaturated long chain fatty acid content of, for example, microbial cells or animals. Host cells are manipulated to express a sense or antisense transcript of a DNA encoding a polypeptide(s) which catalyzes the conversion of DGLA to ARA. The substrate(s) for the expressed enzyme may be produced by the host cell or may be exogenously supplied. To achieve expression, the transformed DNA is

operably associated with transcriptional and translational initiation and termination regulatory regions that are functional in the host cell. Constructs comprising the gene to be expressed can provide for integration into the genome of the host cell or can autonomously replicate in the host cell. For production of ARA, the expression cassettes generally used include a cassette which provides for Δ5-desaturase activity, particularly in a host cell which produces or can take up DGLA. Production of ω6-type unsaturated fatty acids, such as ARA, is favored in a host microorganism or animal which is substantially free of ALA. The host is selected or obtained by removing or inhibiting activity of a Δ15- or ω3- type desaturase (see Figure 2). The endogenous desaturase activity can be affected by providing an expression cassette for an antisense Δ15 or ω3 transcript, by disrupting a target Δ15- or ω3-desaturase gene through insertion, substitution and/or deletion of all or part of the target gene, or by adding a Δ15- or ω3-desaturase inhibitor. Production of LA also can be increased by providing expression cassettes for Δ9 and/or Δ12-desaturases where their respective enzymatic activities are limiting.

MICROBIAL PRODUCTION OF FATTY ACIDS

Microbial production of fatty acids has several advantages over purification from natural sources such as fish or plants. Many microbes are known with greatly simplified oil compositions compared with those of higher organisms, making purification of desired components easier. Microbial production is not subject to fluctuations caused by external variables such as weather and food supply. Microbially produced oil is substantially free of contamination by environmental pollutants. Additionally, microbes can provide PUFAs in particular forms which may have specific uses. For example, *Spirulina* can provide PUFAs predominantly at the first and third positions of triglycerides; digestion by pancreatic lipases preferentially releases fatty acids from these positions. Following human or animal ingestion of triglycerides derived from *Spirulina*, these PUFAs are released by pancreatic lipases as free

fatty acids and thus are directly available, for example, for infant brain development. Additionally, microbial oil production can be manipulated by controlling culture conditions, notably by providing particular substrates for microbially expressed enzymes, or by addition of compounds which suppress undesired biochemical pathways. In addition to these advantages, production of fatty acids from recombinant microbes provides the ability to alter the naturally occurring microbial fatty acid profile by providing new synthetic pathways in the host or by suppressing undesired pathways, thereby increasing levels of desired PUFAs, or conjugated forms thereof, and decreasing levels of undesired PUFAs.

PRODUCTION OF FATTY ACIDS IN ANIMALS

Production of fatty acids in animals also presents several advantages. Expression of desaturase genes in animals can produce greatly increased levels of desired PUFAs in animal tissues, making recovery from those tissues more economical. For example, where the desired PUFAs are expressed in the breast milk of animals, methods of isolating PUFAs from animal milk are well established. In addition to providing a source for purification of desired PUFAs, animal breast milk can be manipulated through expression of desaturase genes, either alone or in combination with other human genes, to provide animal milks with a PUFA composition substantially similar to human breast milk during the different stages of infant development. Humanized animal milks could serve as infant formulas where human nursing is impossible or undesired, or in cases of malnourishment or disease.

Depending upon the host cell, the availability of substrate, and the desired end product(s), several polypeptides, particularly desaturases, are of interest. By "desaturase" is intended a polypeptide which can desaturate one or more fatty acids to produce a mono- or poly-unsaturated fatty acid or precursor thereof of interest. Of particular interest are polypeptides which can catalyze the conversion of DGLA to produce ARA which includes enzymes which desaturate at the $\Delta5$ position. By "polypeptide" is meant any chain of amino acids, regardless of length or post-translational modification, for example,

glycosylation or phosphorylation. Considerations for choosing a specific polypeptide having desaturase activity include the pH optimum of the polypeptide, whether the polypeptide is a rate limiting enzyme or a component thereof, whether the desaturase used is essential for synthesis of a desired polyunsaturated fatty acid, and/or co-factors required by the polypeptide. The expressed polypeptide preferably has parameters compatible with the biochemical environment of its location in the host cell. For example, the polypeptide may have to compete for substrate with other enzymes in the host cell. Analyses of the K_m and specific activity of the polypeptide in question therefore are considered in determining the suitability of a given polypeptide for modifying PUFA production in a given host cell. The polypeptide used in a particular situation is one which can function under the conditions present in the intended host cell but otherwise can be any polypeptide having desaturase activity which has the desired characteristic of being capable of modifying the relative production of a desired PUFA.

For production of ARA, the DNA sequence used encodes a polypeptide having Δ5-desaturase activity. In particular instances, this can be coupled with an expression cassette which provides for production of a polypeptide having Δ6-desaturase activity and the host cell can optionally be depleted of any Δ15-desaturase activity present, for example by providing a transcription cassette for production of antisense sequences to the Δ15-desaturase transcription product, by disrupting the Δ15-desaturase gene, or by using a host cell which naturally has, or has been mutated to have, low Δ15-desaturase activity. Inhibition of undesired desaturase pathways also can be accomplished through the use of specific desaturase inhibitors such as those described in U.S. Patent No. 4,778,630. The choice of combination of cassettes used can depend in part on the PUFA profile of the host cell. Where the host cell Δ5-desaturase activity is limiting, overexpression of Δ5-desaturase alone generally will be sufficient to provide for enhanced ARA production in the presence of an appropriate substrate such as DGLA. ARA production also can be increased by providing expression cassettes for Δ9- or Δ12-desaturase genes when the activities of

those desaturases are limiting. A scheme for the synthesis of arachidonic acid (20:4 $\Delta^{5, 8, 11, 14}$) from palmitic acid (C_{16}) is shown in Figure 1. A key enzyme in this pathway is a Δ5-desaturase which converts DH-γ-linolenic acid (DGLA, eicosatrienoic acid) to ARA. Conversion of α-linolenic acid (ALA) to stearidonic acid by a Δ6-desaturase is also shown. Production of PUFAs in addition to ARA, including EPA and DHA is shown in Figure 2.

SOURCES OF POLYPEPTIDES HAVING DESATURASE ACTIVITY

A source of polypeptides having desaturase activity and oligonucleotides encoding such polypeptides are organisms which produce a desired polyunsaturated fatty acid. As an example, microorganisms having an ability to produce ARA can be used as a source of Δ5-desaturase activity. Such microorganisms include, for example, those belonging to the genera *Mortierella, Conidiobolus, Pythium, Phytophathora, Penicillium, Porphyridium, Coidosporium, Mucor, Fusarium, Aspergillus, Rhodotorula,* and *Entomophthora*. Within the genus *Porphyridium*, of particular interest is *Porphyridium cruentum*. Within the genus *Mortierella*, of particular interest are *Mortierella elongata, Mortierella exigua, Mortierella hygrophila, Mortierella ramanniana*, var. angulispora, and *Mortierella alpina*. Within the genus *Mucor*, of particular interest are *Mucor circinelloides* and *Mucor javanicus*.

DNAs encoding desired desaturases can be identified in a variety of ways. As an example, a source of the desired desaturase, for example genomic or cDNA libraries from *Mortierella*, is screened with detectable enzymatically- or chemically-synthesized probes, which can be made from DNA, RNA, or non-naturally occurring nucleotides, or mixtures thereof. Probes may be enzymatically synthesized from DNAs of known desaturases for normal or reduced-stringency hybridization methods. Oligonucleotide probes also can be used to screen sources and can be based on sequences of known desaturases, including sequences conserved among known desaturases, or on peptide sequences obtained from the desired purified protein. Oligonucleotide probes based on amino acid sequences can be degenerate to encompass the degeneracy

of the genetic code, or can be biased in favor of the preferred codons of the source organism. Oligonucleotides also can be used as primers for PCR from reverse transcribed mRNA from a known or suspected source; the PCR product can be the full length cDNA or can be used to generate a probe to obtain the desired full length cDNA. Alternatively, a desired protein can be entirely sequenced and total synthesis of a DNA encoding that polypeptide performed.

Once the desired genomic or cDNA has been isolated, it can be sequenced by known methods. It is recognized in the art that such methods are subject to errors, such that multiple sequencing of the same region is routine and is still expected to lead to measurable rates of mistakes in the resulting deduced sequence, particularly in regions having repeated domains, extensive secondary structure, or unusual base compositions, such as regions with high GC base content. When discrepancies arise, resequencing can be done and can employ special methods. Special methods can include altering sequencing conditions by using: different temperatures; different enzymes; proteins which alter the ability of oligonucleotides to form higher order structures; altered nucleotides such as ITP or methylated dGTP; different gel compositions, for example adding formamide; different primers or primers located at different distances from the problem region; or different templates such as single stranded DNAs. Sequencing of mRNA also can be employed.

For the most part, some or all of the coding sequence for the polypeptide having desaturase activity is from a natural source. In some situations, however, it is desirable to modify all or a portion of the codons, for example, to enhance expression, by employing host preferred codons. Host preferred codons can be determined from the codons of highest frequency in the proteins expressed in the largest amount in a particular host species of interest. Thus, the coding sequence for a polypeptide having desaturase activity can be synthesized in whole or in part. All or portions of the DNA also can be synthesized to remove any destabilizing sequences or regions of secondary structure which would be present in the transcribed mRNA. All or portions of the DNA also can be synthesized to alter the base composition to one more

preferable in the desired host cell. Methods for synthesizing sequences and bringing sequences together are well established in the literature. *In vitro* mutagenesis and selection, site-directed mutagenesis, or other means can be employed to obtain mutations of naturally occurring desaturase genes to produce a polypeptide having desaturase activity *in vivo* with more desirable physical and kinetic parameters for function in the host cell, such as a longer half-life or a higher rate of production of a desired polyunsaturated fatty acid.

Mortierella alpina **Desaturase**

Of particular interest is the *Mortierella alpina* Δ5-desaturase which has 446 amino acids; the amino acid sequence is shown in Figure 3. The gene encoding the *Mortierella alpina* Δ5-desaturase can be expressed in transgenic microorganisms or animals to effect greater synthesis of ARA from DGLA. Other DNAs which are substantially identical to the *Mortierella alpina* Δ5-desaturase DNA, or which encode polypeptides which are substantially identical to the *Mortierella alpina* Δ5-desaturase polypeptide, also can be used. By substantially identical is intended an amino acid sequence or nucleic acid sequence exhibiting in order of increasing preference at least 60%, 80%, 90% or 95% homology to the *Mortierella alpina* Δ5-desaturase amino acid sequence or nucleic acid sequence encoding the amino acid sequence. For polypeptides, the length of comparison sequences generally is at least 16 amino acids, preferably at least 20 amino acids, or most preferably 35 amino acids. For nucleic acids, the length of comparison sequences generally is at least 50 nucleotides, preferably at least 60 nucleotides, and more preferably at least 75 nucleotides, and most preferably, 110 nucleotides. Homology typically is measured using sequence analysis software, for example, the Sequence Analysis software package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wisconsin 53705, MEGAlign (DNAStar, Inc., 1228 S. Park St., Madison, Wisconsin 53715), and MacVector (Oxford Molecular Group, 2105 S. Bascom Avenue, Suite 200, Campbell, California 95008). Such software matches similar sequences by

assigning degrees of homology to various substitutions, deletions, and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine and leucine; aspartic acid, glutamic acid, asparagine, and glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine. Substitutions may also be made on the basis of conserved hydrophobicity or hydrophilicity (Kyte and Doolittle, *J. Mol. Biol.* 157: 105-132, 1982), or on the basis of the ability to assume similar polypeptide secondary structure (Chou and Fasman, *Adv. Enzymol.* 47: 45-148, 1978).

Other Desaturases

Encompassed by the present invention are related desaturases from the same or other organisms. Such related desaturases include variants of the disclosed Δ5-desaturase naturally occurring within the same or different species of *Mortierella*, as well as homologues of the disclosed Δ5-desaturase from other species. Also included are desaturases which, although not substantially identical to the *Mortierella alpina* Δ5-desaturase, desaturate a fatty acid molecule at carbon 5 from the carboxyl end of a fatty acid molecule. Related desaturases can be identified by their ability to function substantially the same as the disclosed desaturases; that is, are still able to effectively convert DGLA to ARA. Related desaturases also can be identified by screening sequence databases for sequences homologous to the disclosed desaturase, by hybridization of a probe based on the disclosed desaturase to a library constructed from the source organism, or by RT-PCR using mRNA from the source organism and primers based on the disclosed desaturase. Such desaturases include those from humans, *Dictyostelium discoideum* and *Phaeodactylum tricornum.*

The regions of a desaturase polypeptide important for desaturase activity can be determined through routine mutagenesis, expression of the resulting mutant polypeptides and determination of their activities. Mutants may include deletions, insertions and point mutations, or combinations thereof. A typical

functional analysis begins with deletion mutagenesis to determine the N- and C-terminal limits of the protein necessary for function, and then internal deletions, insertions or point mutants are made to further determine regions necessary for function. Other techniques such as cassette mutagenesis or total synthesis also can be used. Deletion mutagenesis is accomplished, for example, by using exonucleases to sequentially remove the 5' or 3' coding regions. Kits are available for such techniques. After deletion, the coding region is completed by ligating oligonucleotides containing start or stop codons to the deleted coding region after 5' or 3' deletion, respectively. Alternatively, oligonucleotides encoding start or stop codons are inserted into the coding region by a variety of methods including site-directed mutagenesis, mutagenic PCR or by ligation onto DNA digested at existing restriction sites. Internal deletions can similarly be made through a variety of methods including the use of existing restriction sites in the DNA, by use of mutagenic primers via site directed mutagenesis or mutagenic PCR. Insertions are made through methods such as linker-scanning mutagenesis, site-directed mutagenesis or mutagenic PCR. Point mutations are made through techniques such as site-directed mutagenesis or mutagenic PCR.

Chemical mutagenesis also can be used for identifying regions of a desaturase polypeptide important for activity. A mutated construct is expressed, and the ability of the resulting altered protein to function as a desaturase is assayed. Such structure-function analysis can determine which regions may be deleted, which regions tolerate insertions, and which point mutations allow the mutant protein to function in substantially the same way as the native desaturase. All such mutant proteins and nucleotide sequences encoding them are within the scope of the present invention.

EXPRESSION OF DESATURASE GENES

Once the DNA encoding a desaturase polypeptide has been obtained, it is placed in a vector capable of replication in a host cell, or is propagated *in vitro* by means of techniques such as PCR or long PCR. Replicating vectors can include plasmids, phage, viruses, cosmids and the like. Desirable vectors include those useful for mutagenesis of the gene of interest or for expression of

the gene of interest in host cells. The technique of long PCR has made *in vitro* propagation of large constructs possible, so that modifications to the gene of interest, such as mutagenesis or addition of expression signals, and propagation of the resulting constructs can occur entirely *in vitro* without the use of a replicating vector or a host cell.

For expression of a desaturase polypeptide, functional transcriptional and translational initiation and termination regions are operably linked to the DNA encoding the desaturase polypeptide. Expression of the polypeptide coding region can take place *in vitro* or in a host cell. Transcriptional and translational initiation and termination regions are derived from a variety of nonexclusive sources, including the DNA to be expressed, genes known or suspected to be capable of expression in the desired system, expression vectors, chemical synthesis, or from an endogenous locus in a host cell.

Expression In Vitro

In vitro expression can be accomplished, for example, by placing the coding region for the desaturase polypeptide in an expression vector designed for *in vitro* use and adding rabbit reticulocyte lysate and cofactors; labeled amino acids can be incorporated if desired. Such *in vitro* expression vectors may provide some or all of the expression signals necessary in the system used. These methods are well known in the art and the components of the system are commercially available. The reaction mixture can then be assayed directly for the polypeptide, for example by determining its activity, or the synthesized polypeptide can be purified and then assayed.

Expression In A Host Cell

Expression in a host cell can be accomplished in a transient or stable fashion. Transient expression can occur from introduced constructs which contain expression signals functional in the host cell, but which constructs do not replicate and rarely integrate in the host cell, or where the host cell is not proliferating. Transient expression also can be accomplished by inducing the

activity of a regulatable promoter operably linked to the gene of interest, although such inducible systems frequently exhibit a low basal level of expression. Stable expression can be achieved by introduction of a construct that can integrate into the host genome or that autonomously replicates in the host cell. Stable expression of the gene of interest can be selected for through the use of a selectable marker located on or transfected with the expression construct, followed by selection for cells expressing the marker. When stable expression results from integration, integration of constructs can occur randomly within the host genome or can be targeted through the use of constructs containing regions of homology with the host genome sufficient to target recombination with the host locus. Where constructs are targeted to an endogenous locus, all or some of the transcriptional and translational regulatory regions can be provided by the endogenous locus.

When increased expression of the desaturase polypeptide in the source organism is desired, several methods can be employed. Additional genes encoding the desaturase polypeptide can be introduced into the host organism. Expression from the native desaturase locus also can be increased through homologous recombination, for example by inserting a stronger promoter into the host genome to cause increased expression, by removing destabilizing sequences from either the mRNA or the encoded protein by deleting that information from the host genome, or by adding stabilizing sequences to the mRNA (USPN 4,910,141).

When it is desirable to express more than one different gene, appropriate regulatory regions and expression methods, introduced genes can be propagated in the host cell through use of replicating vectors or by integration into the host genome. Where two or more genes are expressed from separate replicating vectors, it is desirable that each vector has a different means of replication. Each introduced construct, whether integrated or not, should have a different means of selection and should lack homology to the other constructs to maintain stable expression and prevent reassortment of elements among constructs. Judicious choices of regulatory regions, selection means and method of

propagation of the introduced construct can be experimentally determined so that all introduced genes are expressed at the necessary levels to provide for synthesis of the desired products.

As an example, where the host cell is a yeast, transcriptional and translational regions functional in yeast cells are provided, particularly from the host species. The transcriptional initiation regulatory regions can be obtained, for example from genes in the glycolytic pathway, such as alcohol dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase (GPD), phosphoglucoisomerase, phosphoglycerate kinase, etc. or regulatable genes such as acid phosphatase, lactase, metallothionein, glucoamylase, etc. Any one of a number of regulatory sequences can be used in a particular situation, depending upon whether constitutive or induced transcription is desired, the particular efficiency of the promoter in conjunction with the open-reading frame of interest, the ability to join a strong promoter with a control region from a different promoter which allows for inducible transcription, ease of construction, and the like. Of particular interest are promoters which are activated in the presence of galactose. Galactose-inducible promoters (GAL1, GAL7, and GAL10) have been extensively utilized for high level and regulated expression of protein in yeast (Lue *et al.*, *Mol. Cell. Biol.* Vol. 7, p. 3446, 1987; Johnston, *Microbiol. Rev.* Vol. 51, p. 458, 1987). Transcription from the GAL promoters is activated by the GAL4 protein, which binds to the promoter region and activates transcription when galactose is present. In the absence of galactose, the antagonist GAL80 binds to GAL4 and prevents GAL4 from activating transcription. Addition of galactose prevents GAL80 from inhibiting activation by GAL4.

Nucleotide sequences surrounding the translational initiation codon ATG have been found to affect expression in yeast cells. If the desired polypeptide is poorly expressed in yeast, the nucleotide sequences of exogenous genes can be modified to include an efficient yeast translation initiation sequence to obtain optimal gene expression. For expression in *Saccharomyces*, this can be done by site-directed mutagenesis of an inefficiently expressed gene

by fusing it in-frame to an endogenous *Saccharomyces* gene, preferably a highly expressed gene, such as the lactase gene.

The termination region can be derived from the 3' region of the gene from which the initiation region was obtained or from a different gene. A large number of termination regions are known to and have been found to be satisfactory in a variety of hosts from the same and different genera and species. The termination region usually is selected more as a matter of convenience rather than because of any particular property. Preferably, the termination region is derived from a yeast gene, particularly *Saccharomyces*, *Schizosaccharomyces*, *Candida* or *Kluyveromyces*. The 3' regions of two mammalian genes, γ interferon and $\alpha 2$ interferon, are also known to function in yeast.

INTRODUCTION OF CONSTRUCTS INTO HOST CELLS

Constructs comprising the gene of interest may be introduced into a host cell by standard techniques. These techniques include transformation, protoplast fusion, lipofection, transfection, transduction, conjugation, infection, bolistic impact, electroporation, microinjection, scraping, or any other method which introduces the gene of interest into the host cell. Methods of transformation which are used include lithium acetate transformation (*Methods in Enzymology*, Vol. 194, p. 186-187, 1991). For convenience, a host cell which has been manipulated by any method to take up a DNA sequence or construct will be referred to as "transformed" or "recombinant" herein.

The subject host will have at least have one copy of the expression construct and may have two or more, depending upon whether the gene is integrated into the genome, amplified, or is present on an extrachromosomal element having multiple copy numbers. Where the subject host is a yeast, four principal types of yeast plasmid vectors can be used: Yeast Integrating plasmids (YIps), Yeast Replicating plasmids (YRps), Yeast Centromere plasmids (YCps), and Yeast Episomal plasmids (YEps). YIps lack a yeast replication origin and must be propagated as integrated elements in the yeast genome. YRps have a chromosomally derived autonomously replicating sequence and

are propagated as medium copy number (20 to 40), autonomously replicating, unstably segregating plasmids. YCps have both a replication origin and a centromere sequence and propagate as low copy number (10-20), autonomously replicating, stably segregating plasmids. YEps have an origin of replication from the yeast 2μm plasmid and are propagated as high copy number, autonomously replicating, irregularly segregating plasmids. The presence of the plasmids in yeast can be ensured by maintaining selection for a marker on the plasmid. Of particular interest are the yeast vectors pYES2 (a YEp plasmid available from Invitrogen, confers uracil prototrophy and a GAL1 galactose-inducible promoter for expression), pRS425-pG1 (a YEp plasmid obtained from Dr. T. H. Chang, Ass. Professor of Molecular Genetics, Ohio State University, containing a constitutive GPD promoter and conferring leucine prototrophy), and pYX424 (a YEp plasmid having a constitutive TP1 promoter and conferring leucine prototrophy; Alber, T. and Kawasaki, G. (1982). *J. Mol. & Appl. Genetics* 1: 419).

The transformed host cell can be identified by selection for a marker contained on the introduced construct. Alternatively, a separate marker construct may be introduced with the desired construct, as many transformation techniques introduce many DNA molecules into host cells. Typically, transformed hosts are selected for their ability to grow on selective media. Selective media may incorporate an antibiotic or lack a factor necessary for growth of the untransformed host, such as a nutrient or growth factor. An introduced marker gene therefor may confer antibiotic resistance, or encode an essential growth factor or enzyme, and permit growth on selective media when expressed in the transformed host. Selection of a transformed host also can occur when the expressed marker protein can be detected, either directly or indirectly. The marker protein may be expressed alone or as a fusion to another protein. The marker protein can be detected by its enzymatic activity; for example β galactosidase can convert the substrate X-gal to a colored product, and luciferase can convert luciferin to a light-emitting product. The marker protein can be detected by its light-producing or modifying characteristics; for

example, the green fluorescent protein of *Aequorea victoria* fluoresces when illuminated with blue light. Antibodies can be used to detect the marker protein or a molecular tag on, for example, a protein of interest. Cells expressing the marker protein or tag can be selected, for example, visually, or by techniques such as FACS or panning using antibodies. For selection of yeast transformants, any marker that functions in yeast may be used. Desirably, resistance to kanamycin and the amino glycoside G418 are of interest, as well as ability to grow on media lacking uracil, leucine, lysine or tryptophan.

The Δ5-desaturase-mediated production of PUFAs can be performed in either prokaryotic or eukaryotic host cells. Prokaryotic cells of interest include *Eschericia*, *Bacillus*, *Lactobacillus*, *cyanobacteria* and the like. Eukaryotic cells include mammalian cells such as those of lactating animals, avian cells such as of chickens, and other cells amenable to genetic manipulation including insect, fungal, and algae cells. The cells may be cultured or formed as part or all of a host organism including an animal. Viruses and bacteriophage also may be used with the cells in the production of PUFAs, particularly for gene transfer, cellular targeting and selection. In a preferred embodiment, the host is any microorganism or animal which produces DGLA and/or can assimilate exogenously supplied DGLA, and preferably produces large amounts of DGLA. Examples of host animals include mice, rats, rabbits, chickens, quail, turkeys, bovines, sheep, pigs, goats, yaks, etc., which are amenable to genetic manipulation and cloning for rapid expansion of the transgene expressing population. For animals, a Δ5-desaturase transgene can be adapted for expression in target organelles, tissues and body fluids through modification of the gene regulatory regions. Of particular interest is the production of PUFAs in the breast milk of the host animal.

Expression In Yeast

Examples of host microorganisms include *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, or other yeast such as *Candida, Kluyveromyces* or other fungi, for example, filamentous fungi such as *Aspergillus, Neurospora,*

Penicillium, etc. Desirable characteristics of a host microorganism are, for example, that it is genetically well characterized, can be used for high level expression of the product using ultra-high density fermentation, and is on the GRAS (generally recognized as safe) list since the proposed end product is intended for ingestion by humans. Of particular interest is use of a yeast, more particularly baker's yeast (*S. cerevisiae*), as a cell host in the subject invention. Strains of particular interest are SC334 (Mat α pep4-3 prbl-1122 ura3-52 leu2-3, 112 regl-501 gal1; *Gene* 83:57-64, 1989, Hovland P. *et al.*), YTC34 (α ade2-101 his3Δ200 lys2-801 ura3-52; obtained from Dr. T. H. Chang, Ass. Professor of Molecular Genetics, Ohio State University), YTC41 (a/α ura3-52/ura3=52 lys2-801/lys2-801 ade2-101/ade2-101 trp1-Δ1/trp1-Δ1 his3Δ200/his3Δ200 leu2Δ1/leu2Δ1; obtained from Dr. T. H. Chang, Ass. Professor of Molecular Genetics, Ohio State University), BJ1995 (obtained from the Yeast Genetic Stock Centre, 1021 Donner Laboratory, Berkeley, CA 94720), INVSC1 (Mat α hiw3Δ1 leu2 trp1-289 ura3-52; obtained from Invitrogen, 1600 Faraday Ave., Carlsbad, CA 92008) and INVSC2 (Mat α his3Δ200 ura3-167; obtained from Invitrogen).

Expression In Avian Species

For producing PUFAs in avian species and cells, such as chickens, turkeys, quail and ducks, gene transfer can be performed by introducing a nucleic acid sequence encoding a Δ5-desaturase into the cells following procedures known in the art. If a transgenic animal is desired, pluripotent stem cells of embryos can be provided with a vector carrying a Δ5-desaturase encoding transgene and developed into adult animal (USPN 5,162,215; Ono *et al.* (1996) *Comparative Biochemistry and Physiology A 113*(3):287-292; WO 9612793; WO 9606160). In most cases, the transgene will be modified to express high levels of the desaturase in order to increase production of PUFAs. The transgene can be modified, for example, by providing transcriptional and/or translational regulatory regions that function in avian cells, such as promoters which direct expression in particular tissues and egg parts such as yolk. The

gene regulatory regions can be obtained from a variety of sources, including chicken anemia or avian leukosis viruses or avian genes such as a chicken ovalbumin gene.

Expression In Insect Cells

Production of PUFAs in insect cells can be conducted using baculovirus expression vectors harboring a Δ5-desaturase transgene. Baculovirus expression vectors are available from several commercial sources such as Clonetech. Methods for producing hybrid and transgenic strains of algae, such as marine algae, which contain and express a desaturase transgene also are provided. For example, transgenic marine algae may be prepared as described in USPN 5,426,040. As with the other expression systems described above, the timing, extent of expression and activity of the desaturase transgene can be regulated by fitting the polypeptide coding sequence with the appropriate transcriptional and translational regulatory regions selected for a particular use. Of particular interest are promoter regions which can be induced under preselected growth conditions. For example, introduction of temperature sensitive and/or metabolite responsive mutations into the desaturase transgene coding sequences, its regulatory regions, and/or the genome of cells into which the transgene is introduced can be used for this purpose.

Expression In Plants

Production of PUFA's in plants can be conducted using various plant transformation systems such as the use of *Agrobacterium tumefaciens*, plant viruses, particle cell transformation and the like which are disclosed in Applicant's related applications U.S. Application Serial Nos. 08/834,033 and 08/956,985 and continuation-in-part applications filed simultaneously with this application all of which are hereby incorporated by reference.

The transformed host cell is grown under appropriate conditions adapted for a desired end result. For host cells grown in culture, the conditions are

typically optimized to produce the greatest or most economical yield of PUFAs, which relates to the selected desaturase activity. Media conditions which may be optimized include: carbon source, nitrogen source, addition of substrate, final concentration of added substrate, form of substrate added, aerobic or anaerobic growth, growth temperature, inducing agent, induction temperature, growth phase at induction, growth phase at harvest, pH, density, and maintenance of selection. Microorganisms such as yeast, for example, are preferably grown using selected media of interest, which include yeast peptone broth (YPD) and minimal media (contains amino acids, yeast nitrogen base, and ammonium sulfate, and lacks a component for selection, for example uracil). Desirably, substrates to be added are first dissolved in ethanol. Where necessary, expression of the polypeptide of interest may be induced, for example by including or adding galactose to induce expression from a GAL promoter.

Expression In An Animal

Expression in cells of a host animal can likewise be accomplished in a transient or stable manner. Transient expression can be accomplished via known methods, for example infection or lipofection, and can be repeated in order to maintain desired expression levels of the introduced construct (*see* Ebert, PCT publication WO 94/05782). Stable expression can be accomplished via integration of a construct into the host genome, resulting in a transgenic animal. The construct can be introduced, for example, by microinjection of the construct into the pronuclei of a fertilized egg, or by transfection, retroviral infection or other techniques whereby the construct is introduced into a cell line which may form or be incorporated into an adult animal (U.S. Patent No. 4,873,191; U.S. Patent No. 5,530,177; U.S. Patent No. 5,565,362; U.S. Patent No. 5,366,894; Wilmut *et al.* (1997) *Nature* 385:810). The recombinant eggs or embryos are transferred to a surrogate mother (U.S. Patent No. 4,873,191; U.S. Patent No. 5,530,177; U.S. Patent No. 5,565,362; U.S. Patent No. 5,366,894; Wilmut *et al.* (supra)).

After birth, transgenic animals are identified, for example, by the presence of an introduced marker gene, such as for coat color, or by PCR or Southern blotting from a blood, milk or tissue sample to detect the introduced construct, or by an immunological or enzymological assay to detect the expressed protein or the products produced therefrom (U.S. Patent No. 4,873,191; U.S. Patent No. 5,530,177; U.S. Patent No. 5,565,362; U.S. Patent No. 5,366,894; Wilmut *et al.* (supra)). The resulting transgenic animals may be entirely transgenic or may be mosaics, having the transgenes in only a subset of their cells. The advent of mammalian cloning, accomplished by fusing a nucleated cell with an enucleated egg, followed by transfer into a surrogate mother, presents the possibility of rapid, large-scale production upon obtaining a "founder" animal or cell comprising the introduced construct; prior to this, it was necessary for the transgene to be present in the germ line of the animal for propagation (Wilmut *et al.* (supra)).

Expression in a host animal presents certain efficiencies, particularly where the host is a domesticated animal. For production of PUFAs in a fluid readily obtainable from the host animal, such as milk, the desaturase transgene can be expressed in mammary cells from a female host, and the PUFA content of the host cells altered. The desaturase transgene can be adapted for expression so that it is retained in the mammary cells, or secreted into milk, to form the PUFA reaction products localized to the milk (PCT publication WO 95/24488). Expression can be targeted for expression in mammary tissue using specific regulatory sequences, such as those of bovine α-lactalbumin, α-casein, β-casein, γ-casein, κ-casein, β-lactoglobulin, or whey acidic protein, and may optionally include one or more introns and/or secretory signal sequences (U.S. Patent No. 5,530,177; Rosen, U.S. Patent No. 5,565,362; Clark *et al.*, U.S. Patent No. 5,366,894; Garner *et al.*, PCT publication WO 95/23868). Expression of desaturase transgenes, or antisense desaturase transcripts, adapted in this manner can be used to alter the levels of specific PUFAs, or derivatives thereof, found in the animals milk. Additionally, the Δ5-desaturase transgene can be expressed either by itself or with other transgenes, in order to produce

animal milk containing higher proportions of desired PUFAs or PUFA ratios and concentrations that resemble human breast milk (Prieto *et al.*, PCT publication WO 95/24494).

PURIFICATION OF FATTY ACIDS

The fatty acids desaturated in the Δ5 position may be found in the host microorganism or animal as free fatty acids or in conjugated forms such as acylglycerols, phospholipids, sulfolipids or glycolipids, and may be extracted from the host cell through a variety of means well-known in the art. Such means may include extraction with organic solvents, sonication, supercritical fluid extraction using for example carbon dioxide, and physical means such as presses, or combinations thereof. Of particular interest is extraction with methanol and chloroform. Where desirable, the aqueous layer can be acidified to protonate negatively charged moieties and thereby increase partitioning of desired products into the organic layer. After extraction, the organic solvents can be removed by evaporation under a stream of nitrogen. When isolated in conjugated forms, the products may be enzymatically or chemically cleaved to release the free fatty acid or a less complex conjugate of interest, and can then be subject to further manipulations to produce a desired end product. Desirably, conjugated forms of fatty acids are cleaved with potassium hydroxide.

If further purification is necessary, standard methods can be employed. Such methods may include extraction, treatment with urea, fractional crystallization, HPLC, fractional distillation, silica gel chromatography, high speed centrifugation or distillation, or combinations of these techniques. Protection of reactive groups, such as the acid or alkenyl groups, may be done at any step through known techniques, for example alkylation or iodination. Methods used include methylation of the fatty acids to produce methyl esters. Similarly, protecting groups may be removed at any step. Desirably, purification of fractions containing ARA, DHA and EPA may be accomplished by treatment with urea and/or fractional distillation.

USES OF FATTY ACIDS

There are several uses for fatty acids of the subject invention. Probes based on the DNAs of the present invention may find use in methods for isolating related molecules or in methods to detect organisms expressing desaturases. When used as probes, the DNAs or oligonucleotides must be detectable. This is usually accomplished by attaching a label either at an internal site, for example via incorporation of a modified residue, or at the 5' or 3' terminus. Such labels can be directly detectable, can bind to a secondary molecule that is detectably labeled, or can bind to an unlabelled secondary molecule and a detectably labeled tertiary molecule; this process can be extended as long as is practical to achieve a satisfactorily detectable signal without unacceptable levels of background signal. Secondary, tertiary, or bridging systems can include use of antibodies directed against any other molecule, including labels or other antibodies, or can involve any molecules which bind to each other, for example a biotin-streptavidin/avidin system. Detectable labels typically include radioactive isotopes, molecules which chemically or enzymatically produce or alter light, enzymes which produce detectable reaction products, magnetic molecules, fluorescent molecules or molecules whose fluorescence or light-emitting characteristics change upon binding. Examples of labelling methods can be found in USPN 5,011,770. Alternatively, the binding of target molecules can be directly detected by measuring the change in heat of solution on binding of probe to target via isothermal titration calorimetry, or by coating the probe or target on a surface and detecting the change in scattering of light from the surface produced by binding of target or probe, respectively, as may be done with the BIAcore system.

PUFAs produced by recombinant means find applications in a wide variety of areas. Supplementation of humans or animals with PUFAs in various forms can result in increased levels not only of the added PUFAs, but of their metabolic progeny as well. For example, where the inherent Δ5-desaturase pathway is dysfunctional in an individual, treatment with ARA can result not

only in increased levels of ARA, but also of downstream products of ARA such as prostaglandins (see Figure 1). Complex regulatory mechanisms can make it desirable to combine various PUFAs, or to add different conjugates of PUFAs, in order to prevent, control or overcome such mechanisms to achieve the desired levels of specific PUFAs in an individual.

NUTRITIONAL COMPOSITIONS

The present invention also includes nutritional compositions. Such compositions, for purposes of the present invention, include any food or preparation for human consumption including for enteral or parenteral consumption, which when taken into the body (a) serve to nourish or build up tissues or supply energy and/or (b) maintain, restore or support adequate nutritional status or metabolic function.

The nutritional composition of the present invention comprises at least one oil or acid produced in accordance with the present invention and may either be in a solid or liquid form. Additionally, the composition may include edible macronutrients, vitamins and minerals in amounts desired for a particular use. The amount of such ingredients will vary depending on whether the composition is intended for use with normal, healthy infants, children or adults having specialized needs such as those which accompany certain metabolic conditions (e.g., metabolic disorders).

Examples of macronutrients which may be added to the composition include but are not limited to edible fats, carbohydrates and proteins. Examples of such edible fats include but are not limited to coconut oil, soy oil, and mono- and diglycerides. Examples of such carbohydrates include but are not limited to glucose, edible lactose and hydrolyzed search. Additionally, examples of proteins which may be utilized in the nutritional composition of the invention include but are not limited to soy proteins, electrodialysed whey , electrodialysed skim milk, milk whey, or the hydrolysates of these proteins.

With respect to vitamins and minerals, the following may be added to the nutritional compositions of the present invention: calcium, phosphorus,

potassium, sodium, chloride, magnesium, manganese, iron, copper, zinc, selenium, iodine, and Vitamins A, E, D, C, and the B complex. Other such vitamins and minerals may also be added.

The components utilized in the nutritional compositions of the present invention will of semi-purified or purified origin. By semi-purified or purified is meant a material which has been prepared by purification of a natural material or by synthesis.

Examples of nutritional compositions of the present invention include but are not limited to infant formulas, dietary supplements, and rehydration compositions. Nutritional compositions of particular interest include but are not limited to those utilized for enteral and parenteral supplementation for infants, specialist infant formulae, supplements for the elderly, and supplements for those with gastrointestinal difficulties and/or malabsorption.

Nutritional Compositions

A typical nutritional composition of the present invention will contain edible macronutrients, vitamins and minerals in amounts desired for a particular use. The amounts of such ingredients will vary depending on whether the formulation is intended for use with normal, healthy individuals temporarily exposed to stress, or to subjects having specialized needs due to certain chronic or acute disease states (e.g., metabolic disorders). It will be understood by persons skilled in the art that the components utilized in a nutritional formulation of the present invention are of semi-purified or purified origin. By semi-purified or purified is meant a material that has been prepared by purification of a natural material or by synthesis. These techniques are well known in the art (See, e.g., Code of Federal Regulations for Food Ingredients and Food Processing; Recommended Dietary Allowances, 10th Ed., National Academy Press, Washington, D.C., 1989).

In a preferred embodiment, a nutritional formulation of the present invention is an enteral nutritional product, more preferably an adult or child

enteral nutritional product. Accordingly in a further aspect of the invention, a nutritional formulation is provided that is suitable for feeding adults, who are experiencing stress. The formula comprises, in addition to the PUFAs of the invention; macronutrients, vitamins and minerals in amounts designed to provide the daily nutritional requirements of adults.

The macronutritional components include edible fats, carbohydrates and proteins. Exemplary edible fats are coconut oil, soy oil, and mono- and diglycerides and the PUFA oils of this invention. Exemplary carbohydrates are glucose, edible lactose and hydrolyzed cornstarch. A typical protein source would be soy protein, electrodialysed whey or electrodialysed skim milk or milk whey, or the hydrolysates of these proteins, although other protein sources are also available and may be used. These macronutrients would be added in the form of commonly accepted nutritional compounds in amount equivalent to those present in human milk or an energy basis, i.e., on a per calorie basis.

Methods for formulating liquid and enteral nutritional formulas are well known in the art and are described in detail in the examples.

The enteral formula can be sterilized and subsequently utilized on a ready-to-feed (RTF) basis or stored in a concentrated liquid or a powder. The powder can be prepared by spray drying the enteral formula prepared as indicated above, and the formula can be reconstituted by rehydrating the concentrate. Adult and infant nutritional formulas are well known in the art and commercially available (e.g., Similac®, Ensure®, Jevity® and Alimentum® from Ross Products Division, Abbott Laboratories). An oil or acid of the present invention can be added to any of these formulas in the amounts described below.

The energy density of the nutritional composition when in liquid form, can typically range from about 0.6 Kcal to 3.0 Kcal per ml. When in solid or powdered form, the nutritional supplement can contain from about 1.2 to more than 9 Kcals per gm, preferably 3 to 7 Kcals per gram. In general, the osmolality of a liquid product should be less than 700 mOsm and more preferably less than 660 mOsm.

The nutritional formula would typically include vitamins and minerals, in addition to the PUFAs of the invention, in order to help the individual ingest the minimum daily requirements for these substances. In addition to the PUFAs listed above, it may also be desirable to supplement the nutritional composition with zinc, copper, and folic acid in addition to antioxidants. It is believed that these substances will also provide a boost to the stressed immune system and thus will provide further benefits to the individual. The presence of zinc, copper or folic acid is optional and is not required in order to gain the beneficial effects on immune suppression. Likewise a pharmaceutical composition can be supplemented with these same substances as well.

In a more preferred embodiment, the nutritional contains, in addition to the antioxidant system and the PUFA component, a source of carbohydrate wherein at least 5 weight % of said carbohydrate is an indigestible oligosaccharide.. In yet a more preferred embodiment, the nutritional composition additionally contains protein, taurine and carnitine.

The PUFAs, or derivatives thereof, made by the disclosed method can be used as dietary substitutes, or supplements, particularly infant formulas, for patients undergoing intravenous feeding or for preventing or treating malnutrition. Typically, human breast milk has a fatty acid profile comprising from about 0.15 % to about 0.36 % as DHA, from about 0.03 % to about 0.13 % as EPA, from about 0.30 % to about 0.88 % as ARA, from about 0.22 % to about 0.67 % as DGLA, and from about 0.27 % to about 1.04 % as GLA. Additionally, the predominant triglyceride in human milk has been reported to be 1,3-di-oleoyl-2-palmitoyl, with 2-palmitoyl glycerides reported as better absorbed than 2-oleoyl or 2-lineoyl glycerides (USPN 4,876,107). Thus, fatty acids such as ARA, DGLA, GLA and/or EPA produced by the invention can be used to alter the composition of infant formulas to better replicate the PUFA composition of human breast milk. In particular, an oil composition for use in a pharmacologic or food supplement, particularly a breast milk substitute or supplement, will preferably comprise one or more of ARA, DGLA and GLA.

More preferably the oil will comprise from about 0.3 to 30% ARA, from about 0.2 to 30% DGLA, and from about 0.2 to about 30% GLA.

In addition to the concentration, the ratios of ARA, DGLA and GLA can be adapted for a particular given end use. When formulated as a breast milk supplement, or substitute an oil composition which contains two or more of ARA, DGLA and GLA will be provided in a ratio of about 1:19:30 to about 6:1:0.2, respectively. For example, the breast milk of animals can vary in ratios of ARA:DGLA:DGL ranging from 1:19:30 to 6:1:0.2, which includes intermediate ratios which are preferably about 1:1:1, 1:2:1, 1:1:4. When produced together in a host cell, adjusting the rate and percent of conversion of a precursor substrate such as GLA and DGLA to ARA can be used to precisely control the PUFA ratios. For example, a 5% to 10% conversion rate of DGLA to ARA can be used to produce an ARA to DGLA ratio of about 1:19, whereas a conversion rate of about 75% to 80% can be used to produce an ARA to DGLA ratio of about 6:1. Therefore, whether in a cell culture system or in a host animal, regulating the timing, extent and specificity of desaturase expression as described can be used to modulate the PUFA levels and ratios. Depending on the expression system used, e.g., cell culture or an animal expressing oil(s) in its milk, the oils also can be isolated and recombined in the desired concentrations and ratios. Amounts of oils providing these ratios of PUFA can be determined following standard protocols. PUFAs, or host cells containing them, also can be used as animal food supplements to alter an animal's tissue or milk fatty acid composition to one more desirable for human or animal consumption.

For dietary supplementation, the purified PUFAs, or derivatives thereof, may be incorporated into cooking oils, fats or margarines formulated so that in normal use the recipient would receive the desired amount. The PUFAs may also be incorporated into infant formulas, nutritional supplements or other food products, and may find use as anti-inflammatory or cholesterol lowering agents.

Pharmaceutical Compositi ns

The present invention also encompasses a pharmaceutical composition comprising one or more of the acids and/or resulting oils produced in accordance with the methods described herein. More specifically, such a pharmaceutical composition may comprise one or more of the acids and/or oils as well as a standard, well-known, non-toxic pharmaceutically acceptable carrier, adjuvant or vehicle such as, for example, phosphate buffered saline, water, ethanol, polyols, vegetable oils, a wetting agent or an emulsion such as a water/oil emulsion. The composition may be in either a liquid or solid form. For example, the composition may be in the form of a tablet, capsule, ingestible liquid or powder, injectible, or topical ointment or cream.

Possible routes of administration include, for example, oral, rectal and parenteral. The route of administration will, of course, depend upon the desired effect. For example, if the composition is being utilized to treat rough, dry, or aging skin, to treat injured or burned skin, or to treat skin or hair affected by a disease or condition, it may perhaps be applied topically.

The dosage of the composition to be administered to the patient may be determined by one of ordinary skill in the art and depends upon various factors such as weight of the patient, age of the patient, immune status of the patient, etc.

With respect to form, the composition may be, for example, a solution, a dispersion, a suspension, an emulsion or a sterile powder which is then reconstituted.

Additionally, the composition of the present invention may be utilized for cosmetic purposes. It may be added to pre-existing cosmetic compositions such that a mixture is formed or may be used as a sole composition.

Pharmaceutical compositions may be utilized to administer the PUFA component to an individual. Suitable pharmaceutical compositions may comprise physiologically acceptable sterile aqueous or non-aqueous solutions, dispersions, suspensions or emulsions and sterile powders for reconstitution into

sterile solutions or dispersions for ingestion. Examples of suitable aqueous and non-aqueous carriers, diluents, solvents or vehicles include water, ethanol, polyols (propyleneglycol, polyethyleneglycol, glycerol, and the like), suitable mixtures thereof, vegetable oils (such as olive oil) and injectable organic esters such as ethyl oleate. Proper fluidity can be maintained, for example, by the maintenance of the required particle size in the case of dispersions and by the use of surfactants. It may also be desirable to include isotonic agents, for example sugars, sodium chloride and the like. Besides such inert diluents, the composition can also include adjuvants, such as wetting agents, emulsifying and suspending agents, sweetening, flavoring and perfuming agents.

Suspensions, in addition to the active compounds, may contain suspending agents, as for example, ethoxylated isostearyl alcohols, polyoxyethylene sorbitol and sorbitan esters, microcrystalline cellulose, aluminum metahydroxide, bentonite, agar-agar and tragacanth or mixtures of these substances, and the like.

Solid dosage forms such as tablets and capsules can be prepared using techniques well known in the art. For example, PUFAs of the invention can be tableted with conventional tablet bases such as lactose, sucrose, and cornstarch in combination with binders such as acacia, cornstarch or gelatin, disintegrating agents such as potato starch or alginic acid and a lubricant such as stearic acid or magnesium stearate. Capsules can be prepared by incorporating these excipients into a gelatin capsule along with the antioxidants and the PUFA component. The amount of the antioxidants and PUFA component that should be incorporated into the pharmaceutical formulation should fit within the guidelines discussed above.

As used in this application, the term "treat" refers to either preventing, or reducing the incidence of, the undesired occurrence. For example, to treat immune suppression refers to either preventing the occurrence of this suppression or reducing the amount of such suppression. The terms "patient" and "individual" are being used interchangeably and both refer to an animal. The term "animal" as used in this application refers to any warm-blooded

mammal including, but not limited to, dogs, humans, monkeys, and apes. As used in the application the term "about" refers to an amount varying from the stated range or number by a reasonable amount depending upon the context of use. Any numerical number or range specified in the specification should be considered to be modified by the term about.

"Dose" and "serving" are used interchangeably and refer to the amount of the nutritional or pharmaceutical composition ingested by the patient in a single setting and designed to deliver effective amounts of the antioxidants and the structured triglyceride. As will be readily apparent to those skilled in the art, a single dose or serving of the liquid nutritional powder should supply the amount of antioxidants and PUFAs discussed above. The amount of the dose or serving should be a volume that a typical adult can consume in one sitting. This amount can vary widely depending upon the age, weight, sex or medical condition of the patient. However as a general guideline, a single serving or dose of a liquid nutritional produce should be considered as encompassing a volume from 100 to 600 ml, more preferably from 125 to 500 ml and most preferably from 125 to 300 ml.

The PUFAs of the present invention may also be added to food even when supplementation of the diet is not required. For example, the composition may be added to food of any type including but not limited to margarines, modified butters, cheeses, milk, yogurt, chocolate, candy, snacks, salad oils, cooking oils, cooking fats, meats, fish and beverages.

Pharmaceutical Applications

For pharmaceutical use (human or veterinary), the compositions are generally administered orally but can be administered by any route by which they may be successfully absorbed, e.g., parenterally (i.e. subcutaneously, intramuscularly or intravenously), rectally or vaginally or topically, for example, as a skin ointment or lotion. The PUFAs of the present invention may be administered alone or in combination with a pharmaceutically acceptable carrier or excipient. Where available, gelatin capsules are the preferred form of oral administration. Dietary supplementation as set forth above also can

provide an oral route of administration. The unsaturated acids of the present invention may be administered in conjugated forms, or as salts, esters, amides or prodrugs of the fatty acids. Any pharmaceutically acceptable salt is encompassed by the present invention; especially preferred are the sodium, potassium or lithium salts. Also encompassed are the N-alkylpolyhydroxamine salts, such as N-methyl glucamine, found in PCT publication WO 96/33155. The preferred esters are the ethyl esters. As solid salts, the PUFAs also can be administered in tablet form. For intravenous administration, the PUFAs or derivatives thereof may be incorporated into commercial formulations such as Intralipids. The typical normal adult plasma fatty acid profile comprises 6.64 to 9.46% of ARA, 1.45 to 3.11% of DGLA, and 0.02 to 0.08% of GLA. These PUFAs or their metabolic precursors can be administered, either alone or in mixtures with other PUFAs, to achieve a normal fatty acid profile in a patient. Where desired, the individual components of formulations may be individually provided in kit form, for single or multiple use. A typical dosage of a particular fatty acid is from 0.1 mg to 20 g, or even 100 g daily, and is preferably from 10 mg to 1, 2, 5 or 10 g daily as required, or molar equivalent amounts of derivative forms thereof. Parenteral nutrition compositions comprising from about 2 to about 30 weight percent fatty acids calculated as triglycerides are encompassed by the present invention; preferred is a composition having from about 1 to about 25 weight percent of the total PUFA composition as GLA (USPN 5,196,198). Other vitamins, and particularly fat-soluble vitamins such as vitamin A, D, E and L-carnitine can optionally be included. Where desired, a preservative such as α tocopherol may be added, typically at about 0.1% by weight.

Suitable pharmaceutical compositions may comprise physiologically acceptable sterile aqueous or non-aqueous solutions, dispersions, suspensions or emulsions and sterile powders for reconstitution into sterile injectible solutions or dispersions. Examples of suitable aqueous and non-aqeuous carriers, diluents, solvents or vehicles include water, ethanol, polyols (propylleneglyol, polyethylenegycol, glycerol, and the like), suitable mixtures thereof, vegetable

oils (such as olive oil) and injectable organic esters such as ehyl oleate. Proper fluidity can be maintained, for example, by the maintenance of the required particle size in the case of dispersions and by the use of surfactants. It may also be desirable to include isotonic agents, for example sugars, sodium chloride and the like. Besides such inert diluents, the composition can also include adjuvants, such as wetting agents, emulsifying and suspending agents, sweetening, flavoring and perfuming agents.

Suspensions in addition to the active compounds, may contain suspending agents, as for example, ethoxylated isostearyl alcohols, polyoxyethylene sorbitol and sorbitan esters, microcrystalline cellulose, aluminum metahydroxide, bentonite, agar-agar and tragacanth, or mixtures of these substances and the like.

An especially preferred pharmaceutical composition contains diacetyltartaric acid esters of mono- and diglycerides dissolved in an aqueous medium or solvent. Diacetyltartaric acid esters of mono- and diglycerides have an HLB value of about 9-12 and are significantly more hydrophilic than existing antimicrobial lipids that have HLB values of 2-4. Those existing hydrophobic lipids cannot be formulated into aqueous compositions. As disclosed herein, those lipids can now be solubilized into aqueous media in combination with diacetyltartaric acid esters of mono-and diglycerides. In accordance with this embodiment, diacetyltartaric acid esters of mono- and diglycerides (e.g., DATEM-C12:0) is melted with other active antimicrobial lipids (e.g., 18:2 and 12:0 monoglycerides) and mixed to obtain a homogeneous mixture. Homogeneity allows for increased antimicrobial activity. The mixture can be completely dispersed in water. This is not possible without the addition of diacetyltartaric acid esters of mono- and diglycerides and premixing with other monoglycerides prior to introduction into water. The aqueous composition can then be admixed under sterile conditions with physiologically acceptable diluents, preservatives, buffers or propellants as may be required to form a spray or inhalant.

The present invention also encompasses the treatment of numerous disorders with fatty acids. Supplementation with PUFAs of the present invention can be used to treat restenosis after angioplasty. Symptoms of inflammation, rheumatoid arthritis, and asthma and psoriasis can be treated with the PUFAs of the present invention. Evidence indicates that PUFAs may be involved in calcium metabolism, suggesting that PUFAs of the present invention may be used in the treatment or prevention of osteoporosis and of kidney or urinary tract stones.

The PUFAs of the present invention can be used in the treatment of cancer. Malignant cells have been shown to have altered fatty acid compositions; addition of fatty acids has been shown to slow their growth and cause cell death, and to increase their susceptibility to chemotherapeutic agents. GLA has been shown to cause reexpression on cancer cells of the E-cadherin cellular adhesion molecules, loss of which is associated with aggressive metastasis. Clinical testing of intravenous administration of the water soluble lithium salt of GLA to pancreatic cancer patients produced statistically significant increases in their survival. PUFA supplementation may also be useful for treating cachexia associated with cancer.

The PUFAs of the present invention can also be used to treat diabetes (USPN 4,826,877; Horrobin *et al.*, Am. J. Clin. Nutr. Vol. 57 (Suppl.), 732S-737S). Altered fatty acid metabolism and composition has been demonstrated in diabetic animals. These alterations have been suggested to be involved in some of the long-term complications resulting from diabetes, including retinopathy, neuropathy, nephropathy and reproductive system damage. Primrose oil, which contains GLA, has been shown to prevent and reverse diabetic nerve damage.

The PUFAs of the present invention can be used to treat eczema, reduce blood pressure and improve math scores. Essential fatty acid deficiency has been suggested as being involved in eczema, and studies have shown beneficial effects on eczema from treatment with GLA. GLA has also been shown to reduce increases in blood pressure associated with stress, and to improve

performance on arithmetic tests. GLA and DGLA have been shown to inhibit platelet aggregation, cause vasodilation, lower cholesterol levels and inhibit proliferation of vessel wall smooth muscle and fibrous tissue (Brenner *et al.*, Adv. Exp. Med. Biol. Vol. 83, p. 85-101, 1976). Administration of GLA or DGLA, alone or in combination with EPA, has been shown to reduce or prevent gastro-intestinal bleeding and other side effects caused by non-steroidal anti-inflammatory drugs (USPN 4,666,701). GLA and DGLA have also been shown to prevent or treat endometriosis and premenstrual syndrome (USPN 4,758,592) and to treat myalgic encephalomyelitis and chronic fatigue after viral infections (USPN 5,116,871).

Further uses of the PUFAs of this invention include use in treatment of AIDS, multiple schlerosis, acute respiratory syndrome, hypertension and inflammatory skin disorders. The PUFAs of the inventions also can be used for formulas for general health as well as for geriatric treatments.

Veterinary Applications

It should be noted that the above-described pharmaceutical and nutritional compositions may be utilized in connection with animals, as well as humans, as animals experience many of the same needs and conditions as human. For example, the oil or acids of the present invention may be utilized in animal feed supplements.

The following examples are presented by way of illustration, not of limitation.

Examples

Example 1	Isolation of a Δ5-desaturase Nucleotide Sequence from *Mortierella alpina*
Example 2	Expression of *M. alpina* Δ5-desaturase Clones in Baker's Yeast
Example 3	Initial Optimization of Culture Conditions

Example 4	Distribution of PUFAs in Yeast Lipid Fractions
Example 5	Further Culture Optimization
Example 6	Identification of Homologues to *M. alpina* Δ5 and Δ6 desaturases
Example 7	Identification of *M. alpina* Δ5 and Δ6 homologues in other PUFA-producing organisms
Example 8	Identification of *M. alpina* Δ5 and Δ6 homologues in other PUFA-producing organisms
Example 9	Human Desaturase Sequences
Example 10	Nutritional Compositions

Example 1

Isolation of a Δ5-desaturase Nucleotide Sequence from *Mortierella alpina*

Motierella alpina produces arachidonic acid (ARA, 20:4) from the precursor 20:3 by a Δ5-desaturase. A nucleotide sequence encoding the Δ5-desaturase from *Mortierella alpina* was obtained through PCR amplification using *M. alpina* 1st strand cDNA and degenerate oligonucleotide primers corresponding to amino acid sequences conserved between Δ6-desaturases from *Synechocystis* and *Spirulina*. The procedure used was as follows:

Total RNA was isolated from a 3 day old PUFA-producing culture of *Mortierella alpina* using the protocol of Hoge *et al.* (1982) *Experimental Mycology* 6:225-232. The RNA was used to prepare double-stranded cDNA using BRL's lambda-ZipLox system, following the manufacturer's instructions. Several size fractions of the *M. alpina* cDNA were packaged separately to yield libraries with different average-sized inserts. The "full-length" library contains approximately 3 x 10^6 clones with an average insert size of 1.77 kb. The "sequencing-grade" library contains approximately 6 x 10^5 clones with an average insert size of 1.1 kb.

5μg of total RNA was reverse transcribed using BRL Superscript RTase and the primer TSyn (5'-CCAAGCTTCTGCAGGAGCTCTTTTTTT TTTTTTTT-3'), SEQ ID NO:10. Degenerate oligonucleotides were designed to regions conserved between the two cyanobacterial Δ6-desaturase sequences. The specific primers used were D6DESAT-F3 (SEQ ID NO:8) (5'-CUACUACUACUACAYCAYACOTAYACOAAYAT-3') and D6DESAT-R3 (SEQ ID NO:9) (5'-CAUCAUCAUCAUOGGRAAOARRTGRTG-3'), where Y=C+T, R=A+G, and O=I+C. PCR amplification was carried out in a 25μl volume containing: template derived from 40 ng total RNA, 2 pM each primer, 200 μM each deoxyribonucleotide triphosphate, 60 mM Tris-Cl, pH 8.5, 15 mM $(NH_4)_2SO_4$, 2 mM $MgCl_2$. Samples were subjected to an initial denaturation step of 95 degrees (all temperatures Celsius) for 5 minutes, then held at 72 degrees while 0.2 U of Taq polymerase were added. PCR thermocycling conditions were as follows: 94 degrees for 1 min., 45 degrees for 1.5 min., 72 degrees for 2 min. PCR was continued for 35 cycles. PCR using these primers on the *M. alpina* first-strand cDNA produced a 550 bp reaction product. Comparison of the deduced amino acid sequence of the *M. alpina* PCR fragment SEQ ID NO:3 revealed regions of homology with Δ6-desaturases (*see* Figure 5). However, there was only about 28% identity over the region compared.

The PCR product was used as a probe to isolate corresponding cDNA clones from a *M. alpina* library. The longest cDNA clone, Ma29, was designated pCGN5521 and has been completely sequenced on both strands. The cDNA is contained as a 1481 bp insert in the vector pZL1 (Bethesda Research Laboratories) and, beginning with the first ATG, contains an open reading frame encoding 446 amino acids. The reading frame contains the sequence deduced from the PCR fragment. The sequence of the cDNA insert was found to contain regions of homology to Δ6-desaturases (*see* Figure 5). For example, three conserved "histidine boxes" (that have been observed in membrane-bound desaturases (Okuley *et al.*, (1994) *The Plant Cell* 6:147-158)) were found to be present in the *Mortierella* sequence at amino acid positions

171-175, 207-212, and 387-391 (*see* Figure 3). However, the typical "HXXHH" amino acid motif for the third histidine box for the *Mortierella* desaturase was found to be QXXHH, SEQ ID NO:11-12. Surprisingly, the amino-terminus of the encoded protein, showed significant homology to cytochrome b5 proteins. Thus, the *Mortierella* cDNA clone appears to represent a fusion between a cytochrome b5 and a fatty acid desaturase. Since cytochrome b5 is believed to function as the electron donor for membrane-bound desaturase enzymes, it is possible that the N-terminal cytochrome b5 domain of this desaturase protein is involved in its function. This may be advantageous when expressing the desaturase in heterologous systems for PUFA production.

Example 2

Expression of *M. alpina* Desaturase Clones in Baker's Yeast

Yeast Transformation

Lithium acetate transformation of yeast was performed according to standard protocols (*Methods in Enzymology*, Vol. 194, p. 186-187, 1991). Briefly, yeast were grown in YPD at 30°C. Cells were spun down, resuspended in TE, spun down again, resuspended in TE containing 100 mM lithium acetate, spun down again, and resuspended in TE/lithium acetate. The resuspended yeast were incubated at 30°C for 60 minutes with shaking. Carrier DNA was added, and the yeast were aliquoted into tubes. Transforming DNA was added, and the tubes were incubated for 30 min. at 30°C. PEG solution (35% (w/v) PEG 4000, 100 mM lithium acetate, TE pH7.5) was added followed by a 50 min. incubation at 30°C. A 5 min. heat shock at 42°C was performed, the cells were pelleted, washed with TE, pelleted again and resuspended in TE. The resuspended cells were then plated on selective media.

Desaturase Expression in Transformed Yeast

The cDNA clones from *Mortierella alpina* were screened for desaturase activity in baker's yeast. A canola Δ15-desaturase (obtained by PCR using 1st

strand cDNA from *Brassica napus* cultivar 212/86 seeds using primers based on the published sequence (Arondel *et al. Science* 258:1353-1355)) was used as a positive control. The Δ15-desaturase gene and the gene from cDNA clone Ma29 was inserted into the expression vector pYES2 (Invitrogen), resulting in plasmids pCGR-2 and pCGR-4, respectively. These plasmids were transfected into *S. cerevisiae* yeast strain 334 and expressed after induction with galactose and in the presence of substrates that allowed detection of specific desaturase activity. The control strain was *S. cerevisiae* strain 334 containing the unaltered pYES2 vector. The substrates used, the products produced and the indicated desaturase activity were: DGLA (conversion to ARA would indicate Δ5-desaturase activity), linolenic acid (conversion to GLA would indicate Δ6-desaturase activity; conversion to ALA would indicate Δ15-desaturase activity), oleic acid (an endogenous substrate made by *S. cerevisiae*, conversion to linolenic acid would indicate Δ12-desaturase activity, which *S. cerevisiae* lacks), or ARA (conversion to EPA would indicate Δ17-desaturase activity). The results are provided in Table 1 below. The lipid fractions were extracted as follows: Cultures were grown for 48-52 hours at 15°C. Cells were pelleted by centrifugation, washed once with sterile ddH_20, and repelleted. Pellets were vortexed with methanol; chloroform was added along with tritridecanoin (as an internal standard). The mixtures were incubated for at least one hour at room temperature or at 4°C overnight. The chloroform layer was extracted and filtered through a Whatman filter with one gram of anhydrous sodium sulfate to remove particulates and residual water. The organic solvents were evaporated at 40°C under a stream of nitrogen. The extracted lipids were then derivatized to fatty acid methyl esters (FAME) for gas chromatography analysis (GC) by adding 2 ml of 0.5 N potassium hydroxide in methanol to a closed tube. The samples were heated to 95°C to 100°C for 30 minutes and cooled to room temperature. Approximately 2 ml of 14 % boron trifluoride in methanol was added and the heating repeated. After the extracted lipid mixture cooled, 2 ml of water and 1 ml of hexane were added to extract the FAME for analysis by GC. The percent conversion was calculated by dividing the product produced

by the sum of (the product produced and the substrate added) and then multiplying by 100. To calculate the oleic acid percent conversion, as no substrate was added, the total linolenic acid produced was divided by the sum of (oleic acid and linolenic acid produced), then multiplying by 100.

Table 1

M. alpina Desaturase Expression in Baker's Yeast

CLONE	TYPE OF ENZYME ACTIVITY	% CONVERSION OF SUBSTRATE
pCGR-2	Δ6	0 (18:2 to 18:3ω6)
(canola Δ15	Δ15	16.3 (18:2 to 18:3ω3)
desaturase)	Δ5	2.0 (20:3 to 20:4ω6)
	Δ17	2.8 (20:4 to 20:5ω3)
	Δ12	1.8 (18:1 to 18:2ω6)
pCGR-4	Δ6	0
(*M. alpina*	Δ15	0
Ma29)	Δ5	15.3
	Δ17	0.3
	Δ12	3.3

The Δ15-desaturase control clone exhibited 16.3% conversion of the substrate. The pCGR-4 clone expressing the Ma29 cDNA converted 15.3% of the 20:3 substrate to 20:4ω6, indicating that the gene encodes a Δ5-desaturase. The background (non-specific conversion of substrate) was between 0-3% in these cases. We also found substrate inhibition of the activity by using different concentrations of the substrate. When substrate was added to 100 μM, the percent conversion to product dropped compared to when substrate was added to 25 μM (see below). Additionally, by varying the DGLA substrate concentrations, between about 5 μM to about 200 μM percent conversion of DGLA to ARA ranged from about 5% to 75% with the *M. alpina* Δ5-desaturase.

These data show that desaturases with different substrate specificities can be expressed in a heterologous system and used to produce poly-unsaturated long chain fatty acids.

Table 2 represents fatty acids of interest as a percent of the total lipid extracted from the yeast host *S. cerevisiae* 334 with the indicated plasmid. No glucose was present in the growth media. Affinity gas chromatography was used to separate the respective lipids. GC/MS was employed to verify the identity product(s). The expected product for the *B. napus* Δ15-desaturase, α-linolenic acid, was detected when its substrate, linolenic acid, was added exogenously to the induced yeast culture. This finding demonstrates that yeast expression of a desaturase gene can produce functional enzyme and detectable amounts of product under the current growth conditions. Both exogenously added substrates were taken up by yeast, although slightly less of the longer chain PUFA, dihomo-γ-linolenic acid (20:3), was incorporated into yeast than linolenic acid (18:2) when either was added in free form to the induced yeast cultures. Arachidonic acid was detected as a novel PUFA in yeast when dihomo-γ-linolenic acid was added as the substrate to *S. cerevisiae* 334 (pCGR-4). This identifies pCGR-4 (MA29) as the Δ5-desaturase from *M. alpina*. Prior to this, no isolation and expression of a Δ5-desaturase from any source has been reported.

Table 2

Fatty Acid as a Percentage of Total Lipid Extracted from Yeast

Plasmid in Yeast (enzyme)	18:2 Incorporated	α-18:3 Produced	γ-18:3 Produced	20:3 Incorporated	20:4 Produced	18:1* Present	18:2 Produced
pYES2 (control)	66.9	0	0	58.4	0	4	0
pCGR-2 (Δ15)	60.1	5.7	0	50.4	0	0.7	0
pCGR-4 (Δ5)	67	0	0	32.3	5.8	0.8	0

100 μM substrate added

* 18:1 is an endogenous fatty acid in yeast

Key To Tables
18:1 =oleic acid
18:2 =linolenic acid
α-18:3 =α-linolenic acid
γ-18:3 =γ-linolenic acid
18:4 =stearidonic acid
20:3 =dihomo-γ-linolenic acid
20:4 =arachidonic acid

Example 3
Optimization of Culture Conditions

Table 3A shows the effect of exogenous free fatty acid substrate concentration on yeast uptake and conversion to fatty acid product as a percentage of the total yeast lipid extracted. In all instances, low amounts of exogenous substrate (1-10 μM) resulted in low fatty acid substrate uptake and product formation. Between 25 and 50 μM concentration of free fatty acid in the growth and induction media gave the highest percentage of fatty acid product formed, while the 100 μM concentration and subsequent high uptake into yeast appeared to decrease or inhibit the desaturase activity. The feedback inhibition of high fatty acid substrate concentration was well illustrated when the percent conversion rates of the respective fatty acid substrates to their respective products were compared in Table 3B. In all cases, 100 μM substrate concentration in the growth media decreased the percent conversion to product. The effect of media composition was also evident when glucose was present in the growth media for the Δ5-desaturase, since the percent of substrate uptake was decreased at 25 μM (Table 3A). However, the percent conversion by Δ5-desaturase increased by 18% and the percent product formed remained the same in the presence of glucose in the growth media.

Table 3A

Effect of Added Substrate on the Percentage of Incorporated Substrate and Product Formed in Yeast Extracts

Plasmid in Yeast	pCGR-2 (Δ15)	pCGR-4 (Δ5)
substrate/product	18:2 /α-18:3	20:3/20:4
1 μM sub.	ND	0.5/1.7
10μM sub.	ND	3.3/4
25 μ M sub.	ND	5.1/6.1
25 μM◊ sub.	36.6/7.2◊	9.3/5.4◊
50 μM sub.	53.1/6.5◊	ND
100 μM sub.	60.1/5.7◊	32.3/5.8◊

Table 3B
Effect of Substrate Concentration in Media on the Percent Conversion of Fatty Acid Substrate to Product in Yeast Extracts

Plasmid in Yeast	pCGR-2 (Δ15)	pCGR-4 (Δ5)
substrate/product	18:2 →α-18:3	20:3→20:4
1 μM sub.	ND	77.3
10 μM sub.	ND	54.8
25 μM sub.	ND	54.2
25 μM◊ sub.	16.4	36.7
50 μM sub.	10.9◊	ND
100 μM sub.	8.7◊	15.2◊

◊ no glucose in media

+ Yeast peptone broth (YPD)

* 18:1 is an endogenous yeast lipid

sub. is substrate concentration

ND (not done)

Table 4 shows the amount of fatty acid produced by a recombinant desaturase from induced yeast cultures when different amounts of free fatty acid substrate were used. Fatty acid weight was determined since the total amount of lipid varied dramatically when the growth conditions were changed, such as the presence of glucose in the yeast growth and induction media. To better

determine the conditions when the recombinant desaturase would produce the most PUFA product, the quantity of individual fatty acids were examined. The absence of glucose reduced the amount of arachidonic acid produced by Δ5-desaturase by half. For Δ5-desaturase the amount of total yeast lipid was decreased by almost half in the absence of glucose.

Table 4

Fatty Acid Produced in μg from Yeast Extracts

Plasmid in Yeast (enzyme)	pCGR-4 (Δ5)	pCGR-7 (Δ12)
product	20:4	18:2*
1 μM sub.	8.3	ND
10 μM sub.	19.2	ND
25 μM sub.	31.2	115.7
25 μM ◊ sub.	16.8	39 ◊

◊ no glucose in media

sub. is substrate concentration

ND (not done)

*18:1, the substrate, is an endogenous yeast lipid

Example 4

Distribution of PUFAs in Yeast Lipid Fractions

Table 5 illustrates the uptake of free fatty acids and their new products formed in yeast lipids as distributed in the major lipid fractions. A total lipid extract was prepared as described above. The lipid extract was separated on TLC plates, and the fractions were identified by comparison to standards. The bands were collected by scraping, and internal standards were added. The fractions were then saponified and methylated as above, and subjected to gas chromatography. The gas chromatograph calculated the amount of fatty acid by comparison to a standard. It would appear that the substrates are accessible in the phospholipid form to the desaturases.

Table 5

Fatty Acid Distribution in Various Yeast Lipid Fractions in μg

Fatty acid fraction	Phospholipid	Diglyceride	Free Fatty Acid	Triglyceride	Cholesterol Ester
SC (pCGR-4) substrate 20:3	15.1	1.9	22.9	12.6	3.3
SC (pCGR-4) product 20:4	42.6	0.9	6.8	4.9	0.4

SC = *S. cerevisiae* (plasmid)

Example 5
Further Culture Optimization

The growth and induction conditions for optimal activities of desaturases in *Saccharomyces cerevisiae* were evaluated. Various culture conditions that were manipulated for optimal activity were: I) induction temperature, ii) concentration of inducer, iii) timing of substrate addition, iv) concentration of substance, v) sugar source, vi) growth phase at induction. These studies were done using Δ5-desaturase gene from *Mortierella alpina* (MA 29). In addition, the effect of changing host strain on expression of the Δ5-desaturase gene was also determined.

As described above, the best rate of conversion of substrate to ARA was observed at a substrate concentration of 1 μM, however, the percentage of ARA in the total fatty acids was highest at 25 μM substrate concentration. To determine if the substrate needed to be modified to a readily available form before it could be utilized by the desaturase, the substrate was added either 15 hours before induction or concomitant with inducer addition (indicated as after, in Figure 6A). As it can be seen in Figure 6A, addition of substrate before induction did not have a significant effect on the activity of Δ5-desaturase. In fact, addition of substrate along with the inducer was slightly better for expression/activity of Δ5-desaturase, as ARA levels in the total fatty acids were

higher. However, the rate of conversion of substrate to product was slightly lower.

The effect of inducer concentration of expression/activity of *Mortierella* Δ5-desaturase was examined by inducing SC334/pCGR5 with 0.5 or 2% (w/v) of galactose. As shown in Figures 7A and 7B, expression of Δ5-desaturase was higher when induced with 0.5% galactose. Furthermore, rate of conversion of substrate to product was also better when SC334/pCGR5 was induced with 0.5% galactose vs 2% galactose.

To determine the effect of temperature on Δ5-desaturase activity, the SC334 host strain, transformed with pCGR5 (SC334/pCGR5) was grown and induced at 15° C, 25°C, 30°C and 37°C. The quantity of ARA (20:4n6) produced in SC334/pCGR5 cultures, supplemental with substrate 20:3n6, was measured by fatty acid analysis. Figure 8A depicts the quantity of 20:3n6 and 20:4n6, expressed as percentage of total fatty acids. Figure 8B depicts the rate of conversion of substrate to product. Growth and induction of SC334/pCGR5 at 25°C, was the best for the expression of Δ5-desaturase as evidenced by the highest levels of arachidonic acid in the total fatty acids. Additionally the highest rate of conversion of substrate to product also occurred at 25°C. Growth and induction at 15°C gave the lowest expression of ARA, whereas at 37°C gave the lowest conversion of substrate to product.

The effect of yeast strain on expression of the Δ5-desaturase gene was studied in 5 different host strains; INVSC1, INVSC2, YTC34, YTC41, and SC334, at 15°C and 30°C. At 15°C, SC334 has the highest percentage of ARA in total fatty acids, suggesting higher activity of Δ5-desaturase in SC334. The rate of conversion of substrate to product, however is lowest in SC334 and highest in INVSC1 (Fig. 9A and B). At 30°C, the highest percentage of product (ARA) in total fatty acids was observed in INVSC2, although the rate of conversion of substrate to product in INVSC2 was slightly lower than INVSC1 (Fig. 10A and B).

ARA, the product of Δ5-desaturase, is stored in the phospholipid faction (Example 4). Therefore the quantity of ARA produced in yeast is limited by the amount that can be stored in the phospholipid fraction. If ARA could also be stored in other fractions such as the triglyceride fraction, the quantity of ARA produced in yeast might be increased. To test this hypothesis, the Δ5-desaturase gene was expressed in the yeast host strain DBY746 (obtained from the Yeast Genetic Stock Centre, 1021 Donner Laboratory, Berkeley, CA 94720. The genotype of strain DBY746 is Matα, his3-Δ1, leu2-3, leu2-112, ura3-32, trp1-289, gal). The DBY746 yeast strain has an endogenous gene for choline transferase. The presence of this enzyme might enable the DBY746 strain to convert excess phospholipids into triglycerides fraction. Results in Fig. 11 show no increase in the conversion of substrate to product as compared to SC334, which does not have the gene for choline transferase.

To study the effect of media on expression of Δ5-desaturase, pCGR4/SC334 was grown in four different media at two different temperatures (15°C and 30°) and in two different host strains (SC334 and INVSC1). The composition of the media was as follows:

Media A: mm-Ura, + 2% galactose + 2% glucose.

Media B: mm-Ura, + 20% galactose + 2% Glucose + 1M sorbitol (pH5.8)

Media C: mm-Ura, + 2% galactose + 2% raffinose

Media D: mm-Ura, + 2% galactose +2% raffinose + 1M sorbitol (pH5.8)

mm=minimal media

Results show that the highest conversion rate of substrate to product at 15°C in SC334 was observed in media A. The highest conversion rate overall for Δ5-desaturase in SC334 was at 30° in media D. The highest conversion rate of Δ5-desaturase in INVSC1 was also at 30° in media D (Figures 12A and 12B).

These data show that a DNA encoding a desaturase that can convert DGLA to ARA can be isolated from *Mortierella alpina* and can be expressed in a heterologous system and used to produce poly-unsaturated long chain fatty

acids. Exemplified is the production of ARA from the precursor DGLA by expression of a Δ5-desaturase in yeast.

Example 6

Identification of Homologues to *M. alpina* Δ5 and Δ6 desaturases

A nucleic acid sequence that encodes a putative Δ5 desaturase was identified through a TBLASTN search of the est databases through NCBI using amino acids 100-446 of Ma29 as a query. The truncated portion of the Ma29 sequence was used to avoid picking up homologies based on the cytochrome b5 portion at the N-terminus of the desaturase. The deduced amino acid sequence of an est from *Dictyostelium discoideum* (accession # C25549) shows very significant homology to Ma29 and lesser, but still significant homology to Ma524. The DNA sequence is presented as SEQ ID NO:13. The amino acid sequence is presented as SEQ ID NO:14.

Example 7

Identification of *M. alpina* Δ5 and Δ6 homologues in other PUFA-producing organisms

To look for desaturases involved in PUFA production, a cDNA library was constructed from total RNA isolated from Phaeodactylum tricornutum. A plasmid-based cDNA library was constructed in pSPORT1 (GIBCO-BRL) following manufacturer's instructions using a commercially available kit (GIBCO-BRL). Random cDNA clones were sequenced and nucleic acid sequences that encode putative Δ5 or Δ6 desaturases were identified through BLAST search of the databases and comparison to Ma29 and Ma524 sequences.

One clone was identified from the *Phaeodactylum* library with homology to Ma29 and Ma524; it is called 144-011-B12. The DNA sequence is presented as SEQ ID NO:15. The amino acid sequence is presented as SEQ ID NO:16.

Example 8

Identification of *M. alpina* Δ5 and Δ6 homologues in other PUFA-producing organisms

To look for desaturases involved in PUFA production, a cDNA library was constructed from total RNA isolated from *Schizochytrium* species. A plasmid-based cDNA library was constructed in pSPORT1 (GIBCO-BRL) following manufacturer's instructions using a commercially available kit (GIBCO-BRL). Random cDNA clones were sequenced and nucleic acid sequences that encode putative Δ5 or Δ6 desaturases were identified through BLAST search of the databases and comparison to Ma29 and Ma524 sequences.

One clone was identified from the *Schizochytrium* library with homology to Ma29 and Ma524; it is called 81-23-C7. This clone contains a ~1 kb insert. Partial sequence was obtained from each end of the clone using the universal forward and reverse sequencing primers. The DNA sequence from the forward primer is presented as SEQ ID NO:17. The peptide sequence is presented as SEQ ID NO:18. The DNA sequence from the reverse primer is presented as SEQ ID NO:19. The amino acid sequence from the reverse primer is presented as SEQ ID NO:20.

Example 9

Human Desaturase Gene Sequences

Human desaturase gene sequences potentially involved in long chain polyunsaturated fatty acid biosynthesis were isolated based on homology between the human cDNA sequences and *Mortierella alpina* desaturase gene sequences. The three conserved "histidine boxes" known to be conserved among membrane-bound desaturases were found. As with some other membrane-bound desaturases the final HXXHH histidine box motif was found to be QXXHH. The amino acid sequence of the putative human desaturases exhibited homology to *M. alpina* Δ5, Δ6, Δ9, and Δ12 desaturases.

The *M. alpina* Δ5 desaturase and Δ6 desaturase cDNA sequences were used to search the LifeSeq database of Incyte Pharmaceuticals, Inc., Palo Alto, California 94304. The Δ5 desaturase sequence was divided into fragments; 1) amino acid no. 1-150, 2) amino acid no. 151-300, and 3) amino acid no. 301-446. The Δ6 desaturase sequence was divided into three fragments; 1) amino acid no. 1-150, 2) amino acid no. 151-300, and 3) amino acid no. 301-457. These polypeptide fragments were searched against the database using the "tblastn" algorithm. This alogarithm compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands).

The polypeptide fragments 2 and 3 of *M. alpina* Δ5 and Δ6 have homologies with the CloneID sequences as outlined in Table 6. The CloneID represents an individual sequence from the Incyte LifeSeq database. After the "tblastn" results have been reviewed, Clone Information was searched with the default settings of Stringency of >=50, and Productscore <=100 for different CloneID numbers. The Clone Information Results displayed the information including the ClusterID, CloneID, Library, HitID, Hit Description. When selected, the ClusterID number displayed the clone information of all the clones that belong in that ClusterID. The Assemble command assembles all of the CloneID which comprise the ClusterID. The following default settings were used for GCG (Genetics Computer Group, University of Wisconsin Biotechnology Center, Madison, Wisconsin 53705) Assembly:

Word Size:	7
Minimum Overlap:	14
Stringency:	0.8
Minimum Identity:	14
Maximum Gap:	10
Gap Weight:	8
Length Weight:	2

GCG Assembly Results displayed the contigs generated on the basis of sequence information within the CloneID. A contig is an alignment of DNA sequences based on areas of homology among these sequences. A new sequence (consensus sequence) was generated based on the aligned DNA sequences within a contig. The contig containing the CloneID was identified, and the ambiguous sites of the consensus sequence was edited based on the alignment of the CloneIDs (see SEQ ID NO:21 - SEQ ID NO:25) to generate the best possible sequence. The procedure was repeated for all six CloneID listed in Table 6. This produced five unique contigs. The edited consensus sequences of the 5 contigs were imported into the Sequencher software program (Gene Codes Corporation, Ann Arbor, Michigan 48 105). These consensus sequences were assembled. The contig 2511785 overlaps with contig 3506132, and this new contig was called 2535 (SEQ ID NO:27). The contigs from the Sequencher program were copied into the Sequence Analysis software package of GCG.

Each contig was translated in all six reading frames into protein sequences. The *M. alpina* Δ5 (MA29) and Δ6 (MA524) sequences were compared with each of the translated contigs using the FastA search (a Pearson and Lipman search for similarity between a query sequence and a group of sequences of the same type (nucleic acid or protein)). Homology among these sequences suggest the open reading frames of each contig. The homology among the *M. alpina* Δ5 and Δ6 to contigs 2535 and 3854933 were utilized to create the final contig called 253538a. Figure 13 is the FastA match of the final contig 253538a and MA29, and Figure 14 is the FastA match of the final contig 253538a and MA524. The DNA sequences for the various contigs are presented in SEQ ID NO:21 -SEQ ID NO:27. The various peptide sequences are shown in SEQ ID NO:28 - SEQ ID NO:34.

Although the open reading frame was generated by merging the two contigs, the contig 2535 shows that there is a unique sequence in the beginning of this contig which does not match with the contig 3854933. Therefore, it is

possible that these contigs were generated from independent desaturase like human genes.

The contig 253538a contains an open reading frame encoding 432 amino acids. It starts with Gln (CAG) and ends with the stop codon (TGA). The contig 253538a aligns with both *M. alpina* Δ5 and Δ6 sequences, suggesting that it could be either of the desaturases, as well as other known desaturases which share homology with each other. The individual contigs listed in Table 6, as well as the intermediate contig 2535 and the final contig 253538a can be utilized to isolate the complete genes for human desaturases.

Uses of the human desaturases

These human sequences can be expressed in yeast and plants utilizing the procedures described in the preceding examples. For expression in mammalian cells and transgenic animals, these genes may provide superior codon bias. These human sequences can also be used to identify related desaturase sequences.

Table 6

Sections of the Desaturases	Clone ID from LifeSeq Database	Keyword
151-300 Δ5	3808675	Fatty acid desaturase
301-446 Δ5	354535	Δ6
151-300 Δ6	3448789	Δ6
151-300 Δ6	1362863	Δ6
151-300 Δ6	2394760	Δ6
301-457 Δ6	3350263	Δ6

Example 10

Nutritional Compositions

The PUFAs of the previous examples can be utilized in various nutritional supplements, infant formulations, nutritional substitutes and other nutrition solutions.

I. INFANT FORMULATIONS

A. Is mil® S y Formula with Iron.

Usage: As a beverage for infants, children and adults with an allergy or sensitivity to cow's milk. A feeding for patients with disorders for which lactose should be avoided: lactase deficiency, lactose intolerance and galactosemia.

Features:

- Soy protein isolate to avoid symptoms of cow's-milk-protein allergy or sensitivity
- Lactose-free formulation to avoid lactose-associated diarrhea
- Low osmolaity (240 mOsm/kg water) to reduce risk of osmotic diarrhea.
- Dual carbohydrates (corn syrup and sucrose) designed to enhance carbohydrate absorption and reduce the risk of exceeding the absorptive capacity of the damaged gut.
- 1.8 mg of Iron (as ferrous sulfate) per 100 Calories to help prevent iron deficiency.
- Recommended levels of vitamins and minerals.
- Vegetable oils to provide recommended levels of essential fatty acids.
- Milk-white color, milk-like consistency and pleasant aroma.

Ingredients: (Pareve, ®) 85% water, 4.9% corn syrup, 2.6% sugar (sucrose), 2.1% soy oil, 1.9% soy protein isolate, 1.4% coconut oil, 0.15% calcium citrate, 0.11 % calcium phosphate tribasic, potassium citrate, potassium phosphate monobasic, potassium chloride, mono- and disglycerides, soy lecithin, carrageenan, ascorbic acid, L-methionine, magnesium chloride, potassium phosphate dibasic, sodium chloride, choline chloride, taurine, ferrous sulfate, m-inositol, alpha-tocopheryl acetate, zinc sulfate, L-carnitine, niacinamide, calcium pantothenate, cupric sulfate, vitamin A palmitate, thiamine chloride hydrochloride, riboflavin, pyridoxine hydrochloride, folic

acid, manganese sulfate, potassium iodide, phylloquinone, biotin, sodium selenite, vitamin D_3 and cyanocobalamin.

B. Isomil® DF Soy Formula For Diarrhea.

Usage: As a short-term feeding for the dietary management of diarrhea in infants and toddlers.

Features:

- First infant formula to contain added dietary fiber from soy fiber specifically for diarrhea management.
- Clinically shown to reduce the duration of loose, watery stools during mild to severe diarrhea in infants.
- Nutritionally complete to meet the nutritional needs of the infant.
- Soy protein isolate with added L-methionine meets or exceeds an infant's requirement for all essential amino acids.
- Lactose-free formulation to avoid lactose-associated diarrhea.
- Low osmolality (240 mOsm/kg water) to reduce the risk of osmotic diarrhea.
- Dual carbohydrates (corn syrup and sucrose) designed to enhance carbohydrate absorption and reduce the risk of exceeding the absorptive capacity of the damaged gut.
- Meets or exceeds the vitamin and mineral levels recommended by the Committee on Nutrition of the American Academy of Pediatrics and required by the Infant Formula Act.
- 1.8 mg of iron (as ferrous sulfate) per 100 Calories to help prevent iron deficiency.
- Vegetable oils to provide recommended levels of essential fatty acids.

Ingredients: (Pareve, ®) 86% water, 4.8% corn syrup, 2.5% sugar (sucrose), 2.1% soy oil, 2.0% soy protein isolate, 1.4% coconut oil, 0.77% soy

fiber, 0.12% calcium citrate, 0.11 % calcium phosphate tribasic, 0.10% potassium citrate, potassium chloride, potassium phosphate monobasic, mono- and disglycerides, soy lecithin, carrageenan, magnesium chloride, ascorbic acid, L-methionine, potassium phosphate dibasic, sodium chloride, choline chloride, taurine, ferrous sulfate, m-inositol, alpha-tocopheryl acetate, zinc sulfate, L-carnitine, niacinamide, calcium pantothenate, cupric sulfate, vitamin A palmitate, thiamine chloride hydrochloride, riboflavin, pyridoxine hydrochloride, folic acid, manganese sulfate, potassium iodide, phylloquinone, biotin, sodium selenite, vitamin D_3 and cyanocobalamin.

C. Isomil® SF Sucrose-Free Soy Formula With Iron.

Usage: As a beverage for infants, children and adults with an allergy or sensitivity to cow's-milk protein or an intolerance to sucrose. A feeding for patients with disorders for which lactose and sucrose should be avoided.

Features:

- Soy protein isolate to avoid symptoms of cow's-milk-protein allergy or sensitivity.
- Lactose-free formulation to avoid lactose-associated diarrhea (carbohydrate source is Polycose® Glucose Polymers).
- Sucrose free for the patient who cannot tolerate sucrose.
- Low osmolality (180 mOsm/kg water) to reduce risk of osmotic diarrhea.
- 1.8 mg of iron (as ferrous sulfate) per 100 Calories to help prevent iron deficiency.
- Recommended levels of vitamins and minerals.
- Vegetable oils to provide recommended levels of essential fatty acids.
- Milk-white color, milk-like consistency and pleasant aroma.

Ingredients: (Pareve, ®) 75% water, 11.8% hydrolized cornstarch, 4.1% soy oil, 4.1% soy protein isolate, 2.8% coconut oil, 1.0% modified cornstarch,

0.38% calcium phosphate tribasic, 0.17% potassium citrate, 0.13% potassium chloride, mono- and disglycerides, soy lecithin, magnesium chloride, abscorbic acid, L-methionine, calcium carbonate, sodium chloride, choline chloride, carrageenan, taurine, ferrous sulfate, m-inositol, alpha-tocopheryl acetate, zinc sulfate, L-carnitine, niacinamide, calcium pantothenate, cupric sulfate, vitamin A palmitate, thiamine chloride hydrochloride, riboflavin, pyridoxine hydrochloride, folic acid, manganese sulfate, potassium iodide, phylloquinone, biotin, sodium selenite, vitamin D_3 and cyanocobalamin.

D. Isomil® 20 Soy Formula With Iron Ready To Feed, 20 Cal/fl oz.

Usage: When a soy feeding is desired.

Ingredients: (Pareve, ®) 85% water, 4.9% corn syrup, 2.6% sugar (sucrose), 2.1% soy oil, 1.9% soy protein isolate, 1.4% coconut oil, 0.15% calcium citrate, 0.11% calcium phosphate tribasic, potassium citrate, potassium phosphate monobasic, potassium chloride, mono- and disglycerides, soy lecithin, carrageenan, abscorbic acid, L-methionine, magnesium chloride, potassium phosphate dibasic, sodium chloride, choline chloride, taurine, ferrous sulfate, m-inositol, alpha-tocopheryl acetate, zinc sulfate, L-carnitine, niacinamide, calcium pantothenate, cupric sulfate, vitamin A palmitate, thiamine chloride hydrochloride, riboflavin, pyridoxine hydrochloride, folic acid, manganese sulfate, potassium iodide, phylloquinone, biotin, sodium selenite, vitamin D_3 and cyanocobalamin.

E. Similac® Infant Formula

Usage: When an infant formula is needed: if the decision is made to discontinue breastfeeding before age 1 year, if a supplement to breastfeeding is needed or as a routine feeding if breastfeeding is not adopted.

Features:

- Protein of appropriate quality and quantity for good growth; heat-denatured, which reduces the risk of milk-associated enteric blood loss.

- Fat from a blend of vegetable oils (doubly homogenized), providing essential linoleic acid that is easily absorbed.
- Carbohydrate as lactose in proportion similar to that of human milk.
- Low renal solute load to minimize stress on developing organs.
- Powder, Concentrated Liquid and Ready To Feed forms.

Ingredients: (®-D) Water, nonfat milk, lactose, soy oil, coconut oil, mono- and diglycerides, soy lecithin, abscorbic acid, carrageenan, choline chloride, taurine, m-inositol, alpha-tocopheryl acetate, zinc sulfate, niacinamid, ferrous sulfate, calcium pantothenate, cupric sulfate, vitamin A palmitate, thiamine chloride hydrochloride, riboflavin, pyridoxine hydrochloride, folic acid, manganese sulfate, phylloquinone, biotin, sodium selenite, vitamin D_3 and cyanocobalamin.

F. Similac® NeoCare Premature Infant Formula With Iron

Usage: For premature infants' special nutritional needs after hospital discharge. Similac NeoCare is a nutritionally complete formula developed to provide premature infants with extra calories, protein, vitamins and minerals needed to promote catch-up growth and support development.

Features:

- Reduces the need for caloric and vitamin supplementation. More calories (22 Cal/fl oz) then standard term formulas (20 Cal/fl oz).
- Highly absorbed fat blend, with medium-chain triglycerides (MCT oil) to help meet the special digestive needs of premature infants.
- Higher levels of protein, vitamins and minerals per 100 Calories to extend the nutritional support initiated in-hospital.
- More calcium and phosphorus for improved bone mineralization.

Ingredients: ®-D Corn syrup solids, nonfat milk, lactose, whey protein concentrate, soy oil, high-oleic safflower oil, fractionated coconut oil (medium-chain triglycerides), coconut oil, potassium citrate, calcium phosphate tribasic, calcium carbonate, ascorbic acid, magnesium chloride, potassium chloride,

sodium chloride, taurine, ferrous sulfate, m-inositol, choline chloride, ascorbyl palmitate, L-carnitine, alpha-tocopheryl acetate, zinc sulfate, niacinamide, mixed tocopherols, sodium citrate, calcium pantothenate, cupric sulfate, thiamine chloride hydrochloride, vitamin A palmitate, beta carotene, riboflavin, pyridoxine hydrochloride, folic acid, manganese sulfate, phylloquinone, biotin, sodium selenite, vitamin D_3 and cyanocobalamin.

G. Similac Natural Care Low-Iron Human Milk Fortifier Ready To Use, 24 Cal/fl oz.

Usage: Designed to be mixed with human milk or to be fed alternatively with human milk to low-birth-weight infants.

Ingredients: ®-D Water, nonfat milk, hydrolyzed cornstarch, lactose, fractionated coconut oil (medium-chain triglycerides), whey protein concentrate, soil oil, coconut oil, calcium phosphate tribasic, potassium citrate, magnesium chloride, sodium citrate, ascorbic acid, calcium carbonate, mono- and diglycerides, soy lecithin, carrageenan, choline chloride, m-inositol, taurine, niacinamide, L-carnitine, alpha tocopheryl acetate, zinc sulfate, potassium chloride, calcium pantothenate, ferrous sulfate, cupric sulfate, riboflavin, vitamin A palmitate, thiamine chloride hydrochloride, pyridoxine hydrochloride, biotin, folic acid, manganese sulfate, phylloquinone, vitamin D_3, sodium selenite and cyanocobalamin.

Various PUFAs of this invention can be substituted and/or added to the infant formulae described above and to other infant formulae known to those in the art..

II. NUTRITIONAL FORMULATIONS

A. ENSURE®

Usage: ENSURE is a low-residue liquid food designed primarily as an oral nutritional supplement to be used with or between meals or, in appropriate amounts, as a meal replacement. ENSURE is lactose- and gluten-free, and is suitable for use in modified diets, including low-cholesterol diets. Although it is primarily an oral supplement, it can be fed by tube.

Patient C nditi ns:

- For patients on modified diets
- For elderly patients at nutrition risk
- For patients with involuntary weight loss
- For patients recovering from illness or surgery
- For patients who need a low-residue diet

Ingredients:

®-D Water, Sugar (Sucrose), Maltodextrin (Corn), Calcium and Sodium Caseinates, High-Oleic Safflower Oil, Soy Protein Isolate, Soy Oil, Canola Oil, Potassium Citrate, Calcium Phosphate Tribasic, Sodium Citrate, Magnesium Chloride, Magnesium Phosphate Dibasic, Artificial Flavor, Sodium Chloride, Soy Lecithin, Choline Chloride, Ascorbic Acid, Carrageenan, Zinc Sulfate, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Gellan Gum, Niacinamide, Calcium Pantothenate, Manganese Sulfate, Cupric Sulfate, Vitamin A Palmitate, Thiamine Chloride Hydrochloride, Pyridoxine Hydrochloride, Riboflavin, Folic Acid, Sodium Molybdate, Chromium Chloride, Biotin, Potassium Iodide, Sodium Selenate.

B. ENSURE® BARS

Usage: ENSURE BARS are complete, balanced nutrition for supplemental use between or with meals. They provide a delicious, nutrient-rich alternative to other snacks. ENSURE BARS contain <1 g lactose/bar, and Chocolate Fudge Brownie flavor is gluten-free. (Honey Graham Crunch flavor contains gluten.)

Patient Conditions:

- For patients who need extra calories, protein, vitamins and minerals
- Especially useful for people who do not take in enough calories and nutrients

- For people who have the ability to chew and swallow
- Not to be used by anyone with a peanut allergy or any type of allergy to nuts.

Ingredients:

Honey Graham Crunch -- High-Fructose Corn Syrup, Soy Protein Isolate, Brown Sugar, Honey, Maltodextrin (Corn), Crisp Rice (Milled Rice, Sugar [Sucrose], Salt [Sodium Chloride] and Malt), Oat Bran, Partially Hydrogenated Cottonseed and Soy Oils, Soy Polysaccharide, Glycerine, Whey Protein Concentrate, Polydextrose, Fructose, Calcium Caseinate, Cocoa Powder, Artificial Flafors, Canola Oil, High-Oleic Safflower Oil, Nonfat Dry Milk, Whey Powder, Soy Lecithin and Corn Oil. Manufactured in a facility that processes nuts.

Vitamins and Minerals:

Calcium Phosphate Tribasic, Potassium Phosphate Dibasic, Magnesium Oxide, Salt (Sodium Chloride), Potassium Chloride, Ascorbic Acid, Ferric Orthophosphate, Alpha-Tocopheryl Acetate, Niacinamide, Zinc Oxide, Calcium Pantothenate, Copper Gluconate, Manganese Sulfate, Riboflavin, Beta-Carotene, Pyridoxine Hydrochloride, Thiamine Mononitrate, Folic Acid, Biotin, Chromium Chloride, Potassium Iodide, Sodium Selenate, Sodium Molybdate, Phylloquinone, Vitamin D_3 and Cyanocobalamin.

Protein:

Honey Graham Crunch - The protein source is a blend of soy protein isolate and milk proteins.

Soy protein isolate	74%
Milk proteins	26%

Fat:

Honey Graham Crunch - The fat source is a blend of partially hydrogenated cottonseed and soybean, canola, high oleic safflower, and corn oils, and soy lecithin.

Partially hydrogenated cottonseed and soybean oil	76%
Canola oil	8%
High-oleic safflower oil	8%
Corn oil	4%
Soy lecithin	4%

Carbohydrate:

Honey Graham Crunch - The carbohydrate source is a combination of high-fructose corn syrup, brown sugar, maltodextrin, honey, crisp rice, glycerine, soy polysaccharide, and oat bran.

High-fructose corn syrup	24%
Brown sugar	21%
Maltodextrin	12%
Honey	11%
Crisp rice	9%
Glycerine	9%
Soy polysaccharide	7%
Oat bran	7%\

C. ENSURE® HIGH PROTEIN

Usage: ENSURE HIGH PROTEIN is a concentrated, high-protein liquid food designed for people who require additional calories, protein, vitamins, and minerals in their diets. It can be used as an oral nutritional supplement with or between meals or, in appropriate amounts, as a meal replacement. ENSURE HIGH PROTEIN is lactose- and gluten-free, and is suitable for use by people recovering from general surgery or hip fractures and by patients at risk for pressure ulcers.

Patient Conditions

- For patients who require additional calories, protein, vitamins, and minerals, such as patients recovering from general surgery or hip fractures, patients at risk for pressure ulcers, and patients on low-cholesterol diets

Features-

- Low in saturated fat
- Contains 6 g of total fat and < 5 mg of cholesterol per serving
- Rich, creamy taste
- Excellent source of protein, calcium, and other essential vitamins and minerals
- For low-cholesterol diets
- Lactose-free, easily digested

Ingredients:

Vanilla Supreme: -®-D Water, Sugar (Sucrose), Maltodextrin (Corn), Calcium and Sodium Caseinates, High-Oleic Safflower Oil, Soy Protein Isolate, Soy Oil, Canola Oil, Potassium Citrate, Calcium Phosphate Tribasic, Sodium Citrate, Magnesium Chloride, Magnesium Phosphate Dibasic, Artificial Flavor, Sodium Chloride, Soy Lecithin, Choline Chloride, Ascorbic Acid, Carrageenan, Zinc Sulfate, Ferrous Suffate, Alpha-Tocopheryl Acetate, Gellan Gum, Niacinamide, Calcium Pantothenate, Manganese Sulfate, Cupric Sulfate, Vitamin A Palmitate, Thiamine Chloride Hydrochloride, Pyridoxine Hydrochloride, Riboflavin, Folio Acid, Sodium Motybdate, Chromium Chloride, Biotin, Potassium Iodide, Sodium Selenate, Phylloquinone, Vitamin D.3 and Cyanocobalamin.

Protein:

The protein source is a blend of two high-biologic-value proteins: casein and soy.

Sodium and calcium caseinates	85%
Soy protein isolate	15%

Fat:

The fat source is a blend of three oils: high-oleic safflower, canola, and soy.

High-oleic safflower oil	40%

Canola oil	30%
Soy oil	30%

The level of fat in ENSURE HIGH PROTEIN meets American Heart Association (AHA) guidelines. The 6 grams of fat in ENSURE HIGH PROTEIN represent 24% of the total calories, with 2.6% of the fat being from saturated fatty acids and 7.9% from polyunsaturated fatty acids. These values are within the AHA guidelines of $\leq$ 30% of total calories from fat, < 1 0% of the calories from saturated fatty acids, and $\leq$ 1 0% of total calories from polyunsaturated fatty acids.

Carbohydrate:

ENSURE HIGH PROTEIN contains a combination of maltodextrin and sucrose. The mild sweetness and flavor variety (vanilla supreme, chocolate royal, wild berry, and banana), plus VARI-FLAVORSO® Flavor Pacs in pecan, cherry, strawberry, lemon, and orange, help to prevent flavor fatigue and aid in patient compliance.

Vanilla and other nonchocolate flavors

Sucrose	60%
Maltodextrin	40%

Chocolate

Sucrose	70%
Maltodextrin	30%

D. ENSURE ® LIGHT

Usage: ENSURE LIGHT is a low-fat liquid food designed for use as an oral nutritional supplement with or between meals. ENSURE LIGHT is lactose- and gluten-free, and is suitable for use in modified diets, including low-cholesterol diets.

Patient Conditions:

- For normal-weight or overweight patients who need extra nutrition in a supplement that contains 50% less fat and 20% fewer calories than ENSURE
- For healthy adults who don't eat right and need extra nutrition

Features:

- Low in fat and saturated fat
- Contains 3 g of total fat per serving and < 5 mg cholesterol
- Rich, creamy taste
- Excellent source of calcium and other essential vitamins and minerals
- For low-cholesterol diets
- Lactose-free, easily digested

Ingredients:

French Vanilla: ®-D Water, Maltodextrin (Corn), Sugar (Sucrose), Calcium Caseinate, High-Oleic Safflower Oil, Canola Oil, Magnesium Chloride, Sodium Citrate, Potassium Citrate, Potassium Phosphate Dibasic, Magnesium Phosphate Dibasic, Natural and Artificial Flavor, Calcium Phosphate Tribasic, Cellulose Gel, Choline Chloride, Soy Lecithin, Carrageenan, Salt (Sodium Chloride), Ascorbic Acid, Cellulose Gum, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Zinc Sulfate, Niacinamide, Manganese Sulfate, Calcium Pantothenate, Cupric Sulfate, Thiamine Chloride Hydrochloride, Vitamin A Palmitate, Pyridoxine Hydrochloride, Riboflavin, Chromium Chloride, Folic Acid, Sodium Molybdate, Biotin, Potassium Iodide, Sodium Selenate, Phylloquinone, Vitamin D_3 and Cyanocobalamin.

Protein:

The protein source is calcium caseinate.

Calcium caseinate	100%

Fat

The fat source is a blend of two oils: high-oleic safflower and canola.

High-oleic safflower oil	70%
Canola oil	30%

The level of fat in ENSURE LIGHT meets American Heart Association (AHA) guidelines. The 3 grams of fat in ENSURE LIGHT represent 13.5% of the total calories, with 1.4% of the fat being from saturated fatty acids and 2.6% from polyunsaturated fatty acids. These values are within the AHA guidelines of $\leq$ 30% of total calories from fat, < 1 0% of the calories from saturated fatty acids, and $\leq$ 1 0% of total calories from polyunsaturated fatty acids.

Carbohydrate

ENSURE LIGHT contains a combination of maltodextrin and sucrose. The chocolate flavor contains corn syrup as well. The mild sweetness and flavor variety (French vanilla, chocolate supreme, strawberry swirl), plus VARI-FLAVORS® Flavor Pacs in pecan, cherry, strawberry, lemon, and orange, help to prevent flavor fatigue and aid in patient compliance.

Vanilla and other nonchocolate flavors

Sucrose	51%
Maltodextrin	49%

Chocolate

Sucrose	47.0%
Corn Syrup	26.5%
Maltodextrin	26.5%

Vitamins and Minerals

An 8-fl-oz serving of ENSURE LIGHT provides at least 25% of the RDIs for 24 key vitamins and minerals.

Caffeine

Chocolate flavor contains 2.1 mg caffeine/8 fl oz.

E. ENSURE PLUS®

Usage: ENSURE PLUS is a high-calorie, low-residue liquid food for use when extra calories and nutrients, but a normal concentration of protein, are needed. It is designed primarily as an oral nutritional supplement to be used with or between meals or, in appropriate amounts, as a meal replacement. ENSURE PLUS is lactose- and gluten-free. Although it is primarily an oral nutritional supplement, it can be fed by tube.

Patient Conditions:

- For patients who require extra calories and nutrients, but a normal concentration of protein, in a limited volume
- For patients who need to gain or maintain healthy weight

Features

- Rich, creamy taste
- Good source of essential vitamins and minerals

Ingredients

Vanilla: ®-D Water, Corn Syrup, Maltodextrin (Corn), Corn Oil, Sodium and Calcium Caseinates, Sugar (Sucrose), Soy Protein Isolate, Magnesium Chloride, Potassium Citrate, Calcium Phosphate Tribasic, Soy Lecithin, Natural and Artificial Flavor, Sodium Citrate, Potassium Chloride, Choline Chloride, Ascorbic Acid, Carrageenan, Zinc Sulfate, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Niacinamide, Calcium Pantothenate, Manganese Sulfate, Cupric Sulfate, Thiamine Chloride Hydrochloride, Pyridoxine Hydrochloride, Riboflavin, Vitamin A Palmitate, Folic Acid, Biotin, Chromium Chloride, Sodium Molybdate, Potassium Iodide, Sodium Selenite, Phylloquinone, Cyanocobalamin and Vitamin D_3.

Protein

The protein source is a blend of two high-biologic-value proteins: casein and soy.

Sodium and calcium caseinates	84%

Soy protein isolate	16%

Fat

The fat source is corn oil.

Corn oil	100%

Carbohydrate

ENSURE PLUS contains a combination of maltodextrin and sucrose. The mild sweetness and flavor variety (vanilla, chocolate, strawberry. coffee, buffer pecan, and eggnog), plus VARI-FLAVORS® Flavor Pacs in pecan, cherry, strawberry. lemon, and orange, help to prevent flavor fatigue and aid in patient compliance.

Vanilla, strawberry, butter pecan, and coffee flavors

Corn Syrup	39%
Maltodextrin	38%
Sucrose	23%

Chocolate and eggnog flavors

Corn Syrup	36%
Maltodextrin	34%
Sucrose	30%

Vitamins and Minerals

An 8-fl-oz serving of ENSURE PLUS provides at least 15% of the RDIs for 25 key Vitamins and minerals.

Caffeine

Chocolate flavor contains 3.1 mg Caffeine/8 fl oz. Coffee flavor contains a trace amount of caffeine.

F. ENSURE PLUS® HN

Usage: ENSURE PLUS HN is a nutritionally complete high-calorie, high-nitrogen liquid food designed for people with higher calorie and protein needs or limited volume tolerance. It may be used for oral supplementation or for total nutritional support by tube. ENSURE PLUS HN is lactose- and gluten-free.

Patient Conditions:

- For patients with increased calorie and protein needs, such as following surgery or injury
- For patients with limited volume tolerance and early satiety

Features

- For supplemental or total nutrition
- For oral or tube feeding
- 1.5 CaVmL
- High nitrogen
- Calorically dense

Ingredients

Vanilla: ®-D Water, Maltodextrin (Corn), Sodium and Calcium Caseinates, Corn Oil, Sugar (Sucrose), Soy Protein Isolate, Magnesium Chloride, Potassium Citrate, Calcium Phosphate Tribasic, Soy Lecithin, Natural and Artificial Flavor, Sodium Citrate, Choline Chloride, Ascorbic Acid, Taurine, L-Carnitine, Zinc Sulfate, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Niacinamide, Carrageenan, Calcium Pantothenate, Manganese Sulfate, Cupric Sulfate, Thiamine Chloride Hydrochloride, Pyridoxine Hydrochloride, Riboflavin, Vitamin A Palmitate, Folic Acid, Biotin, Chromium Chloride, Sodium Molybdate, Potassium Iodide, Sodium Selenite, Phylloquinone, Cyanocobalamin and Vitamin D_3.

G. ENSURE® POWDER

Usage: ENSURE POWDER (reconstituted with water) is a low-residue liquid food designed primarily as an oral nutritional supplement to be used with or between meals. ENSURE POWDER is lactose- and gluten-free, and is suitable for use in modified diets, including low-cholesterol diets.

Patient Conditions:

- For patients on modified diets
- For elderly patients at nutrition risk
- For patients recovering from illness/surgery
- For patients who need a low-residue diet

Features

- Convenient, easy to mix
- Low in saturated fat
- Contains 9 g of total fat and < 5 mg of cholesterol per serving
- High in vitamins and minerals
- For low-cholesterol diets
- Lactose-free, easily digested

Ingredients: ®-D Corn Syrup, Maltodextrin (Corn), Sugar (Sucrose), Corn Oil, Sodium and Calcium Caseinates, Soy Protein Isolate, Artificial Flavor, Potassium Citrate, Magnesium Chloride, Sodium Citrate, Calcium Phosphate Tribasic, Potassium Chloride, Soy Lecithin, Ascorbic Acid, Choline Chloride, Zinc Sulfate, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Niacinamide, Calcium Pantothenate, Manganese Sulfate, Thiamine Chloride Hydrochloride, Cupric Sulfate, Pyridoxine Hydrochloride, Riboflavin, Vitamin A Palmitate, Folic Acid, Biotin, Sodium Molybdate, Chromium Chloride, Potassium Iodide, Sodium Selenate, Phylloquinone, Vitamin D_3 and Cyanocobalamin.

Protein

The protein source is a blend of two high-biologic-value proteins: casein and soy.

Sodium and calcium caseinates	84%
Soy protein isolate	16%

Fat

The fat source is corn oil.

Corn oil	100%

Carbohydrate

ENSURE POWDER contains a combination of corn syrup, maltodextrin, and sucrose. The mild sweetness of ENSURE POWDER, plus VARI-FLAVORS® Flavor Pacs in pecan, cherry, strawberry, lemon, and orange, helps to prevent flavor fatigue and aid in patient compliance.

Vanilla

Corn Syrup	35%
Maltodextrin	35%
Sucrose	30%

H. ENSURE® PUDDING

Usage: ENSURE PUDDING is a nutrient-dense supplement providing balanced nutrition in a nonliquid form to be used with or between meals. It is appropriate for consistency-modified diets (e.g., soft, pureed, or full liquid) or for people with swallowing impairments. ENSURE PUDDING is gluten-free.

Patient Conditions:

- For patients on consistency-modified diets (e.g., soft, pureed, or full liquid)
- For patients with swallowing impairments
- **Features**
- Rich and creamy, good taste
- Good source of essential vitamins and minerals Convenient-needs no refrigeration

- Gluten-free

Nutrient Profile per 5 oz: Calories 250, Protein 10.9%, Total Fat 34.9%, Carbohydrate 54.2%

Ingredients:

Vanilla: ®-D Nonfat Milk, Water, Sugar (Sucrose), Partially Hydrogenated Soybean Oil, Modified Food Starch, Magnesium Sulfate. Sodium Stearoyl Lactylate, Sodium Phosphate Dibasic, Artificial Flavor, Ascorbic Acid, Zinc Sulfate, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Choline Chloride, Niacinamide, Manganese Sulfate, Calcium Pantothenate, FD&C Yellow #5, Potassium Citrate, Cupric Sulfate, Vitamin A Palmitate, Thiamine Chloride Hydrochloride, Pyridoxine Hydrochloride, Riboflavin, FD&C Yellow #6, Folic Acid, Biotin, Phylloquinone, Vitamin D3 and Cyanocobalamin.

Protein

The protein source is nonfat milk.

Nonfat milk	100%

Fat

The fat source is hydrogenated soybean oil.

Hydrogenated soybean oil	100%

Carbohydrate

ENSURE PUDDING contains a combination of sucrose and modified food starch. The mild sweetness and flavor variety (vanilla, chocolate, butterscotch, and tapioca) help prevent flavor fatigue. The product contains 9.2 grams of lactose per serving.

Vanilla and other nonchocolate flavors

Sucrose	56%
Lactose	27%
Modified food starch	17%

Chocolate

Sucrose	58%
Lactose	26%
Modified food starch	16%

I. ENSURE® WITH FIBER

Usage: ENSURE WITH FIBER is a fiber-containing, nutritionally complete liquid food designed for people who can benefit from increased dietary fiber and nutrients. ENSURE WITH FIBER is suitable for people who do not require a low-residue diet. It can be fed orally or by tube, and can be used as a nutritional supplement to a regular diet or, in appropriate amounts, as a meal replacement. ENSURE WITH FIBER is lactose- and gluten-free, and is suitable for use in modified diets, including low-cholesterol diets.

Patient Conditions

- For patients who can benefit from increased dietary fiber and nutrients

Features

- New advanced formula-low in saturated fat, higher in vitamins and minerals
- Contains 6 g of total fat and < 5 mg of cholesterol per serving
- Rich, creamy taste
- Good source of fiber
- Excellent source of essential vitamins and minerals
- For low-cholesterol diets
- Lactose- and gluten-free

Ingredients

Vanilla: ®-D Water, Maltodextrin (Corn), Sugar (Sucrose), Sodium and Calcium Caseinates, Oat Fiber, High-Oleic Safflower Oil, Canola Oil, Soy Protein Isolate, Corn Oil, Soy Fiber, Calcium Phosphate Tribasic, Magnesium Chloride, Potassium Citrate, Cellulose Gel, Soy Lecithin, Potassium Phosphate

Dibasic, Sodium Citrate, Natural and Artificial Flavors, Choline Chloride, Magnesium Phosphate, Ascorbic Acid, Cellulose Gum, Potassium Chloride, Carrageenan, Ferrous Sulfate, Alpha-Tocopheryl Acetate, Zinc Sulfate, Niacinamide, Manganese Sulfate, Calcium Pantothenate, Cupric Sulfate, Vitamin A Palmitate, Thiamine Chloride Hydrochloride, Pyridoxine Hydrochloride, Riboflavin, Folic Acid, Chromium Chloride, Biotin, Sodium Molybdate, Potassium Iodide, Sodium Selenate, Phylloquinone, Vitamin D_3 and Cyanocobalamin.

Protein

The protein source is a blend of two high-biologic-value proteins- casein and soy.

Sodium and calcium caseinates	80%
Soy protein isolate	20%

Fat

The fat source is a blend of three oils: high-oleic safflower, canola, and corn.

High-oleic safflower oil	40%
Canola oil	40%
Corn oil	20%

The level of fat in ENSURE WITH FIBER meets American Heart Association (AHA) guidelines. The 6 grams of fat in ENSURE WITH FIBER represent 22% of the total calories, with 2.01 % of the fat being from saturated fatty acids and 6.7% from polyunsaturated fatty acids. These values are within the AHA guidelines of $\leq$ 30% of total calories from fat, < 1 0% of the calories from saturated fatty acids, and $\leq$ 1 0% of total calories from polyunsaturated fatty acids.

Carbohydrate

ENSURE WITH FIBER contains a combination of maltodextrin and sucrose. The mild sweetness and flavor variety (vanilla, chocolate, and butter

pecan), plus VARI-FLAVORS® Flavor Pacs in pecan, cherry, strawberry, lemon, and orange, help to prevent flavor fatigue and aid in patient compliance.

Vanilla and other nonchocolate flavors

Maltodextrin	66%
Sucrose	25%
Oat Fiber	7%
Soy Fiber	2%

Chocolate

Maltodextrin	55%
Sucrose	36%
Oat Fiber	7%
Soy Fiber	2%

Fiber

The fiber blend used in ENSURE WITH FIBER consists of oat fiber and soy polysaccharide. This blend results in approximately 4 grams of total dietary fiber per 8-fl-oz can. The ratio of insoluble to soluble fiber is 95:5.

The various nutritional supplements described above and known to others of skill in the art can be substituted and/or supplemented with the PUFAs of this invention.

J. Oxepa™ Nutritional Product

Oxepa is low-carbohydrate, calorically dense enteral nutritional product designed for the dietary management of patients with or at risk for ARDS. It has a unique combination of ingredients, including a patented oil blend containing eicosapentaenoic acid (EPA from fish oil), γ-linolenic acid (GLA from borage oil), and elevated antioxidant levels.

Caloric Distribution:

- Caloric density is high at 1.5 Cal/mL (355 Cal/8 fl oz), to minimize the volume required to meet energy needs.
- The distribution of Calories in Oxepa is shown in Table 7.

Table 7. Caloric Distribution of Oxepa			
	per 8 fl oz.	per liter	% of Cal
Calories	355	1,500	---
Fat (g)	22.2	93.7	55.2
Carbohydrate (g)	25	105.5	28.1
Protein (g)	14.8	62.5	16.7
Water (g)	186	785	---

Fat:

- Oxepa contains 22.2 g of fat per 8-fl oz serving (93.7 g/L).
- The fat source is a oil blend of 31.8% canola oil, 25% medium-chain triglycerides (MCTs), 20% borage oil, 20% fish oil, and 3.2 % soy lecithin. The typical fatty acid profile of Oxepa is shown in Table 8.
- Oxepa provides a balanced amount of polyunsaturated, monounsaturated, and saturated fatty acids, as shown in Table 10.
- Medium-chain trigylcerides (MCTs) -- 25% of the fat blend -- aid gastric emptying because they are absorbed by the intestinal tract without emulsification by bile acids.

The various fatty acid components of Oxepa™ nutritional product can be substituted and/or supplemented with the PUFAs of this invention.

Table 8. Typical Fatty Acid Profile			
	% Total Fatty Acids	g/8 fl oz*	g/L*
Caproic (6:0)	0.2	0.04	0.18
Caprylic (8:0)	14.69	3.1	13.07
Capric (10:0)	11.06	2.33	9.87
Palmitic (16:0)	5.59	1.18	4.98
Palmitoleic (16:1n-7)	1.82	0.38	1.62
Stearic (18:0)	1.84	0.39	1.64

Oleic (18:1n-9)	24.44	5.16	21.75
Linoleic (18:2n-6)	16.28	3.44	14.49
α-Linolenic (18:3n-3)	3.47	0.73	3.09
γ-Linolenic (18:3n-6)	4.82	1.02	4.29
Eicosapentaenoic (20:5n-3)	5.11	1.08	4.55
n-3-Docosapentaenoic (22:5n-3)	0.55	0.12	0.49
Docosahexaenoic (22:6n-3)	2.27	0.48	2.02
Others	7.55	1.52	6.72

* Fatty acids equal approximately 95% of total fat.

Table 9. Fat Profile of Oxepa.	
% of total calories from fat	55.2
Polyunsaturated fatty acids	31.44 g/L
Monounsaturated fatty acids	25.53 g/L
Saturated fatty acids	32.38 g/L
n-6 to n-3 ratio	1.75:1
Cholesterol	9.49 mg/8 fl oz 40.1 mg/L

Carbohydrate:

- The carbohydrate content is 25.0 g per 8-fl-oz serving (105.5 g/L).
- The carbohydrate sources are 45% maltodextrin (a complex carbohydrate) and 55% sucrose (a simple sugar), both of which are readily digested and absorbed.
- The high-fat and low-carbohydrate content of Oxepa is designed to minimize carbon dioxide (CO_2) production. High CO_2 levels can complicate weaning in ventilator-dependent patients. The low level of carbohydrate also may be useful for those patients who have developed stress-induced hyperglycemia.
- Oxepa is lactose-free.

Dietary carbohydrate, the amino acids from protein, and the glycerol moiety of fats can be converted to glucose within the body. Throughout this process, the carbohydrate requirements of glucose-dependent tissues (such as the central nervous system and red blood cells) are met. However, a diet free of

carbohydrates can lead to ketosis, excessive catabolism of tissue protein, and loss of fluid and electrolytes. These effects can be prevented by daily ingestion of 50 to 100 g of digestible carbohydrate, if caloric intake is adequate. The carbohydrate level in Oxepa is also sufficient to minimize gluconeogenesis, if energy needs are being met.

Protein:

- Oxepa contains 14.8 g of protein per 8-fl-oz serving (62.5 g/L).
- The total calorie/nitrogen ratio (150:1) meets the need of stressed patients.
- Oxepa provides enough protein to promote anabolism and the maintenance of lean body mass without precipitating respiratory problems. High protein intakes are a concern in patients with respiratory insufficiency. Although protein has little effect on CO_2 production, a high protein diet will increase ventilatory drive.
- The protein sources of Oxepa are 86.8% sodium caseinate and 13.2% calcium caseinate.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated as incorporated by reference.

The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: KNUTZON, DEBORAH
MURKERJI, PRADIP
HUANG, YUNG-SHENG
THURMOND, JENNIFER
CHAUDHARY, SUNITA
LEONARD, AMANDA

(ii) TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SYNTHESIS OF LONG CHAIN POLY-UNSATURATED FATTY ACIDS

(iii) NUMBER OF SEQUENCES: 34

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: LIMBACH & LIMBACH LLP
(B) STREET: 2001 FERRY BUILDING
(C) CITY: SAN FRANCISCO
(D) STATE: CALIFORNIA
(E) COUNTRY: USA
(F) ZIP: 94111

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: MICHAEL R. WARD
(B) REGISTRATION NUMBER: 38,651
(C) REFERENCE/DOCKET NUMBER: CGAB-110

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (415) 433-4150
(B) TELEFAX: (415) 433-8716
(C) TELEX: N/A

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1483 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GCTTCCTCCA GTTCATCCTC CATTTCGCCA CCTGCATTCT TTACGACCGT TAAGCAAGAT 60

GGGAACGGAC CAAGGAAAAA CCTTCACCTG GGAAGAGCTG GCGGCCCATA ACACCAAGGA 120
CGACCTACTC TTGGCCATCC GCGGCAGGGT GTACGATGTC ACAAAGTTCT TGAGCCGCCA 180
TCCTGGTGGA GTGGACACTC TCCTGCTCGG AGCTGGCCGA GATGTTACTC CGGTCTTTGA 240
GATGTATCAC GCGTTTGGGG CTGCAGATGC CATTATGAAG AAGTACTATG TCGGTACACT 300
GGTCTCGAAT GAGCTGCCCA TCTTCCCGGA GCCAACGGTG TTCCACAAAA CCATCAAGAC 360
GAGAGTCGAG GGCTACTTTA CGGATCGGAA CATTGATCCC AAGAATAGAC CAGAGATCTG 420
GGGACGATAC GCTCTTATCT TTGGATCCTT GATCGCTTCC TACTACGCGC AGCTCTTTGT 480
GCCTTTCGTT GTCGAACGCA CATGGCTTCA GGTGGTGTTT GCAATCATCA TGGGATTTGC 540
GTGCGCACAA GTCGGACTCA ACCCTCTTCA TGATGCGTCT CACTTTTCAG TGACCCACAA 600
CCCCACTGTC TGGAAGATTC TGGGAGCCAC GCACGACTTT TTCAACGGAG CATCGTACCT 660
GGTGTGGATG TACCAACATA TGCTCGGCCA TCACCCCTAC ACCAACATTG CTGGAGCAGA 720
TCCCGACGTG TCGACGTCTG AGCCCGATGT TCGTCGTATC AAGCCCAACC AAAAGTGGTT 780
TGTCAACCAC ATCAACCAGC ACATGTTTGT TCCTTTCCTG TACGGACTGC TGGCGTTCAA 840
GGTGCGCATT CAGGACATCA ACATTTTGTA CTTTGTCAAG ACCAATGACG CTATTCGTGT 900
CAATCCCATC TCGACATGGC ACACTGTGAT GTTCTGGGGC GGCAAGGCTT TCTTTGTCTG 960
GTATCGCCTG ATTGTTCCCC TGCAGTATCT GCCCCTGGGC AAGGTGCTGC TCTTGTTCAC 1020
GGTCGCGGAC ATGGTGTCGT CTTACTGGCT GGCGCTGACC TTCCAGGCGA ACCACGTTGT 1080
TGAGGAAGTT CAGTGGCCGT TGCCTGACGA GAACGGGATC ATCCAAAAGG ACTGGGCAGC 1140
TATGCAGGTC GAGACTACGC AGGATTACGC ACACGATTCG CACCTCTGGA CCAGCATCAC 1200
TGGCAGCTTG AACTACCAGG CTGTGCACCA TCTGTTCCCC AACGTGTCGC AGCACCATTA 1260
TCCCGATATT CTGGCCATCA TCAAGAACAC CTGCAGCGAG TACAAGGTTC CATACCTTGT 1320
CAAGGATACG TTTTGGCAAG CATTTGCTTC ACATTTGGAG CACTTGCGTG TTCTTGGACT 1380
CCGTCCCAAG GAAGAGTAGA AGAAAAAAAG CGCCGAATGA AGTATTGCCC CCTTTTTCTC 1440
CAAGAATGGC AAAAGGAGAT CAAGTGGACA TTCTCTATGA AGA 1483

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 446 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Gly Thr Asp Gln Gly Lys Thr Phe Thr Trp Glu Glu Leu Ala Ala

1 5 10 15

His Asn Thr Lys Asp Asp Leu Leu Leu Ala Ile Arg Gly Arg Val Tyr
20 25 30

Asp Val Thr Lys Phe Leu Ser Arg His Pro Gly Gly Val Asp Thr Leu
35 40 45

Leu Leu Gly Ala Gly Arg Asp Val Thr Pro Val Phe Glu Met Tyr His
50 55 60

Ala Phe Gly Ala Ala Asp Ala Ile Met Lys Lys Tyr Tyr Val Gly Thr
65 70 75 80

Leu Val Ser Asn Glu Leu Pro Ile Phe Pro Glu Pro Thr Val Phe His
85 90 95

Lys Thr Ile Lys Thr Arg Val Glu Gly Tyr Phe Thr Asp Arg Asn Ile
100 105 110

Asp Pro Lys Asn Arg Pro Glu Ile Trp Gly Arg Tyr Ala Leu Ile Phe
115 120 125

Gly Ser Leu Ile Ala Ser Tyr Tyr Ala Gln Leu Phe Val Pro Phe Val
130 135 140

Val Glu Arg Thr Trp Leu Gln Val Val Phe Ala Ile Ile Met Gly Phe
145 150 155 160

Ala Cys Ala Gln Val Gly Leu Asn Pro Leu His Asp Ala Ser His Phe
165 170 175

Ser Val Thr His Asn Pro Thr Val Trp Lys Ile Leu Gly Ala Thr His
180 185 190

Asp Phe Phe Asn Gly Ala Ser Tyr Leu Val Trp Met Tyr Gln His Met
195 200 205

Leu Gly His His Pro Tyr Thr Asn Ile Ala Gly Ala Asp Pro Asp Val
210 215 220

Ser Thr Ser Glu Pro Asp Val Arg Arg Ile Lys Pro Asn Gln Lys Trp
225 230 235 240

Phe Val Asn His Ile Asn Gln His Met Phe Val Pro Phe Leu Tyr Gly
245 250 255

Leu Leu Ala Phe Lys Val Arg Ile Gln Asp Ile Asn Ile Leu Tyr Phe
260 265 270

Val Lys Thr Asn Asp Ala Ile Arg Val Asn Pro Ile Ser Thr Trp His
275 280 285

Thr Val Met Phe Trp Gly Gly Lys Ala Phe Phe Val Trp Tyr Arg Leu
290 295 300

Ile Val Pro Leu Gln Tyr Leu Pro Leu Gly Lys Val Leu Leu Leu Phe
305 310 315 320

Thr Val Ala Asp Met Val Ser Ser Tyr Trp Leu Ala Leu Thr Phe Gln
325 330 335

Ala Asn His Val Val Glu Glu Val Gln Trp Pro Leu Pro Asp Glu Asn
340 345 350

Gly Ile Ile Gln Lys Asp Trp Ala Ala Met Gln Val Glu Thr Thr Gln
355 360 365

Asp Tyr Ala His Asp Ser His Leu Trp Thr Ser Ile Thr Gly Ser Leu
370 375 380

Asn Tyr Gln Ala Val His His Leu Phe Pro Asn Val Ser Gln His His
385 390 395 400

Tyr Pro Asp Ile Leu Ala Ile Ile Lys Asn Thr Cys Ser Glu Tyr Lys
405 410 415

Val Pro Tyr Leu Val Lys Asp Thr Phe Trp Gln Ala Phe Ala Ser His
420 425 430

Leu Glu His Leu Arg Val Leu Gly Leu Arg Pro Lys Glu Glu
435 440 445

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 186 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Leu His His Thr Tyr Thr Asn Ile Ala Gly Ala Asp Pro Asp Val Ser
1 5 10 15

Thr Ser Glu Pro Asp Val Arg Arg Ile Lys Pro Asn Gln Lys Trp Phe
20 25 30

Val Asn His Ile Asn Gln His Met Phe Val Pro Phe Leu Tyr Gly Leu
35 40 45

Leu Ala Phe Lys Val Arg Ile Gln Asp Ile Asn Ile Leu Tyr Phe Val
50 55 60

Lys Thr Asn Asp Ala Ile Arg Val Asn Pro Ile Ser Thr Trp His Thr
65 70 75 80

Val Met Phe Trp Gly Gly Lys Ala Phe Phe Val Trp Tyr Arg Leu Ile
85 90 95

Val Pro Leu Gln Tyr Leu Pro Leu Gly Lys Val Leu Leu Leu Phe Thr
100 105 110

Val Ala Asp Met Val Ser Ser Tyr Trp Leu Ala Leu Thr Phe Gln Ala
115 120 125

Asn Tyr Val Val Glu Glu Val Gln Trp Pro Leu Pro Asp Glu Asn Gly
130 135 140

Ile Ile Gln Lys Asp Trp Ala Ala Met Gln Val Glu Thr Thr Gln Asp

145 150 155 160

Tyr Ala His Asp Ser His Leu Trp Thr Ser Ile Thr Gly Ser Leu Asn
165 170 175

Tyr Gln Xaa Val His His Leu Phe Pro His
180 185

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 457 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ala Ala Ala Pro Ser Val Arg Thr Phe Thr Arg Ala Glu Val Leu
1 5 10 15

Asn Ala Glu Ala Leu Asn Glu Gly Lys Lys Asp Ala Glu Ala Pro Phe
20 25 30

Leu Met Ile Ile Asp Asn Lys Val Tyr Asp Val Arg Glu Phe Val Pro
35 40 45

Asp His Pro Gly Gly Ser Val Ile Leu Thr His Val Gly Lys Asp Gly
50 55 60

Thr Asp Val Phe Asp Thr Phe His Pro Glu Ala Ala Trp Glu Thr Leu
65 70 75 80

Ala Asn Phe Tyr Val Gly Asp Ile Asp Glu Ser Asp Arg Asp Ile Lys
85 90 95

Asn Asp Asp Phe Ala Ala Glu Val Arg Lys Leu Arg Thr Leu Phe Gln
100 105 110

Ser Leu Gly Tyr Tyr Asp Ser Ser Lys Ala Tyr Tyr Ala Phe Lys Val
115 120 125

Ser Phe Asn Leu Cys Ile Trp Gly Leu Ser Thr Val Ile Val Ala Lys
130 135 140

Trp Gly Gln Thr Ser Thr Leu Ala Asn Val Leu Ser Ala Ala Leu Leu
145 150 155 160

Gly Leu Phe Trp Gln Gln Cys Gly Trp Leu Ala His Asp Phe Leu His
165 170 175

His Gln Val Phe Gln Asp Arg Phe Trp Gly Asp Leu Phe Gly Ala Phe
180 185 190

Leu Gly Gly Val Cys Gln Gly Phe Ser Ser Ser Trp Trp Lys Asp Lys
195 200 205

His Asn Thr His His Ala Ala Pro Asn Val His Val Glu Asp Pro Asp
210 215 220

Ile Asp Thr His Pro Leu Leu Thr Trp Ser Glu His Ala Leu Glu Met
225 230 235 240

Phe Ser Asp Val Pro Asp Glu Glu Leu Thr Arg Met Trp Ser Arg Phe
245 250 255

Met Val Leu Asn Gln Thr Trp Phe Tyr Phe Pro Ile Leu Ser Phe Ala
260 265 270

Arg Leu Ser Trp Cys Leu Gln Ser Ile Leu Phe Val Leu Pro Asn Gly
275 280 285

Gln Ala His Lys Pro Ser Gly Ala Arg Val Pro Ile Ser Leu Val Glu
290 295 300

Gln Leu Ser Leu Ala Met His Trp Thr Trp Tyr Leu Ala Thr Met Phe
305 310 315 320

Leu Phe Ile Lys Asp Pro Val Asn Met Leu Val Tyr Phe Leu Val Ser
325 330 335

Gln Ala Val Cys Gly Asn Leu Leu Ala Ile Val Phe Ser Leu Asn His
340 345 350

Asn Gly Met Pro Val Ile Ser Lys Glu Glu Ala Val Asp Met Asp Phe
355 360 365

Phe Thr Lys Gln Ile Ile Thr Gly Arg Asp Val His Pro Gly Leu Phe
370 375 380

Ala Asn Trp Phe Thr Gly Gly Leu Asn Tyr Gln Ile Glu His His Leu
385 390 395 400

Phe Pro Ser Met Pro Arg His Asn Phe Ser Lys Ile Gln Pro Ala Val
405 410 415

Glu Thr Leu Cys Lys Lys Tyr Asn Val Arg Tyr His Thr Thr Gly Met
420 425 430

Ile Glu Gly Thr Ala Glu Val Phe Ser Arg Leu Asn Glu Val Ser Lys
435 440 445

Ala Ala Ser Lys Met Gly Lys Ala Gln
450 455

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 446 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Ala Ala Gln Ile Lys Lys Tyr Ile Thr Ser Asp Glu Leu Lys Asn
1 5 10 15

His Asp Lys Pro Gly Asp Leu Trp Ile Ser Ile Gln Gly Lys Ala Tyr
20 25 30

Asp Val Ser Asp Trp Val Lys Asp His Pro Gly Gly Ser Phe Pro Leu
35 40 45

Lys Ser Leu Ala Gly Gln Glu Val Thr Asp Ala Phe Val Ala Phe His
50 55 60

Pro Ala Ser Thr Trp Lys Asn Leu Asp Lys Phe Phe Thr Gly Tyr Tyr
65 70 75 80

Leu Lys Asp Tyr Ser Val Ser Glu Val Ser Lys Val Tyr Arg Lys Leu
85 90 95

Val Phe Glu Phe Ser Lys Met Gly Leu Tyr Asp Lys Lys Gly His Ile
100 105 110

Met Phe Ala Thr Leu Cys Phe Ile Ala Met Leu Phe Ala Met Ser Val
115 120 125

Tyr Gly Val Leu Phe Cys Glu Gly Val Leu Val His Leu Phe Ser Gly
130 135 140

Cys Leu Met Gly Phe Leu Trp Ile Gln Ser Gly Trp Ile Gly His Asp
145 150 155 160

Ala Gly His Tyr Met Val Val Ser Asp Ser Arg Leu Asn Lys Phe Met
165 170 175

Gly Ile Phe Ala Ala Asn Cys Leu Ser Gly Ile Ser Ile Gly Trp Trp
180 185 190

Lys Trp Asn His Asn Ala His His Ile Ala Cys Asn Ser Leu Glu Tyr
195 200 205

Asp Pro Asp Leu Gln Tyr Ile Pro Phe Leu Val Val Ser Ser Lys Phe
210 215 220

Phe Gly Ser Leu Thr Ser His Phe Tyr Glu Lys Arg Leu Thr Phe Asp
225 230 235 240

Ser Leu Ser Arg Phe Phe Val Ser Tyr Gln His Trp Thr Phe Tyr Pro
245 250 255

Ile Met Cys Ala Ala Arg Leu Asn Met Tyr Val Gln Ser Leu Ile Met
260 265 270

Leu Leu Thr Lys Arg Asn Val Ser Tyr Arg Ala Gln Glu Leu Leu Gly
275 280 285

Cys Leu Val Phe Ser Ile Trp Tyr Pro Leu Leu Val Ser Cys Leu Pro
290 295 300

Asn Trp Gly Glu Arg Ile Met Phe Val Ile Ala Ser Leu Ser Val Thr
305 310 315 320

Gly Met Gln Gln Val Gln Phe Ser Leu Asn His Phe Ser Ser Ser Val
325 330 335

Tyr Val Gly Lys Pro Lys Gly Asn Asn Trp Phe Glu Lys Gln Thr Asp

```
                340                      345                      350

Gly Thr Leu Asp Ile Ser Cys Pro Pro Trp Met Asp Trp Phe His Gly
        355                      360                      365

Gly Leu Gln Phe Gln Ile Glu His His Leu Phe Pro Lys Met Pro Arg
    370                      375                      380

Cys Asn Leu Arg Lys Ile Ser Pro Tyr Val Ile Glu Leu Cys Lys Lys
385                      390                      395                      400

His Asn Leu Pro Tyr Asn Tyr Ala Ser Phe Ser Lys Ala Asn Glu Met
                405                      410                      415

Thr Leu Arg Thr Leu Arg Asn Thr Ala Leu Gln Ala Arg Asp Ile Thr
            420                      425                      430

Lys Pro Leu Pro Lys Asn Leu Val Trp Glu Ala Leu His Thr
        435                      440                      445
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 359 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Leu Thr Ala Glu Arg Ile Lys Phe Thr Gln Lys Arg Gly Phe Arg
1               5                   10                  15

Arg Val Leu Asn Gln Arg Val Asp Ala Tyr Phe Ala Glu His Gly Leu
            20                  25                  30

Thr Gln Arg Asp Asn Pro Ser Met Tyr Leu Lys Thr Leu Ile Ile Val
        35                  40                  45

Leu Trp Leu Phe Ser Ala Trp Ala Phe Val Leu Phe Ala Pro Val Ile
    50                  55                  60

Phe Pro Val Arg Leu Leu Gly Cys Met Val Leu Ala Ile Ala Leu Ala
65                  70                  75                  80

Ala Phe Ser Phe Asn Val Gly His Asp Ala Asn His Asn Ala Tyr Ser
                85                  90                  95

Ser Asn Pro His Ile Asn Arg Val Leu Gly Met Thr Tyr Asp Phe Val
            100                 105                 110

Gly Leu Ser Ser Phe Leu Trp Arg Tyr Arg His Asn Tyr Leu His His
        115                 120                 125

Thr Tyr Thr Asn Ile Leu Gly His Asp Val Glu Ile His Gly Asp Gly
    130                 135                 140

Ala Val Arg Met Ser Pro Glu Gln Glu His Val Gly Ile Tyr Arg Phe
145                 150                 155                 160
```

Gln Gln Phe Tyr Ile Trp Gly Leu Tyr Leu Phe Ile Pro Phe Tyr Trp
165 170 175

Phe Leu Tyr Asp Val Tyr Leu Val Leu Asn Lys Gly Lys Tyr His Asp
180 185 190

His Lys Ile Pro Pro Phe Gln Pro Leu Glu Leu Ala Ser Leu Leu Gly
195 200 205

Ile Lys Leu Leu Trp Leu Gly Tyr Val Phe Gly Leu Pro Leu Ala Leu
210 215 220

Gly Phe Ser Ile Pro Glu Val Leu Ile Gly Ala Ser Val Thr Tyr Met
225 230 235 240

Thr Tyr Gly Ile Val Val Cys Thr Ile Phe Met Leu Ala His Val Leu
245 250 255

Glu Ser Thr Glu Phe Leu Thr Pro Asp Gly Glu Ser Gly Ala Ile Asp
260 265 270

Asp Glu Trp Ala Ile Cys Gln Ile Arg Thr Thr Ala Asn Phe Ala Thr
275 280 285

Asn Asn Pro Phe Trp Asn Trp Phe Cys Gly Gly Leu Asn His Gln Val
290 295 300

Thr His His Leu Phe Pro Asn Ile Cys His Ile His Tyr Pro Gln Leu
305 310 315 320

Glu Asn Ile Ile Lys Asp Val Cys Gln Glu Phe Gly Val Glu Tyr Lys
325 330 335

Val Tyr Pro Thr Phe Lys Ala Ala Ile Ala Ser Asn Tyr Arg Trp Leu
340 345 350

Glu Ala Met Gly Lys Ala Ser
355

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 365 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Thr Ser Thr Thr Ser Lys Val Thr Phe Gly Lys Ser Ile Gly Phe
1 5 10 15

Arg Lys Glu Leu Asn Arg Arg Val Asn Ala Tyr Leu Glu Ala Glu Asn
20 25 30

Ile Ser Pro Arg Asp Asn Pro Pro Met Tyr Leu Lys Thr Ala Ile Ile
35 40 45

```
Leu Ala Trp Val Val Ser Ala Trp Thr Phe Val Val Phe Gly Pro Asp
    50                  55                  60

Val Leu Trp Met Lys Leu Leu Gly Cys Ile Val Leu Gly Phe Gly Val
65                  70                  75                  80

Ser Ala Val Gly Phe Asn Ile Ser His Asp Gly Asn His Gly Gly Tyr
                85                  90                  95

Ser Lys Tyr Gln Trp Val Asn Tyr Leu Ser Gly Leu Thr His Asp Ala
            100                 105                 110

Ile Gly Val Ser Ser Tyr Leu Trp Lys Phe Arg His Asn Val Leu His
        115                 120                 125

His Thr Tyr Thr Asn Ile Leu Gly His Asp Val Glu Ile His Gly Asp
    130                 135                 140

Glu Leu Val Arg Met Ser Pro Ser Met Glu Tyr Arg Trp Tyr His Arg
145                 150                 155                 160

Tyr Gln His Trp Phe Ile Trp Phe Val Tyr Pro Phe Ile Pro Tyr Tyr
                165                 170                 175

Trp Ser Ile Ala Asp Val Gln Thr Met Leu Phe Lys Arg Gln Tyr His
            180                 185                 190

Asp His Glu Ile Pro Ser Pro Thr Trp Val Asp Ile Ala Thr Leu Leu
        195                 200                 205

Ala Phe Lys Ala Phe Gly Val Ala Val Phe Leu Ile Ile Pro Ile Ala
    210                 215                 220

Val Gly Tyr Ser Pro Leu Glu Ala Val Ile Gly Ala Ser Ile Val Tyr
225                 230                 235                 240

Met Thr His Gly Leu Val Ala Cys Val Val Phe Met Leu Ala His Val
                245                 250                 255

Ile Glu Pro Ala Glu Phe Leu Asp Pro Asp Asn Leu His Ile Asp Asp
            260                 265                 270

Glu Trp Ala Ile Ala Gln Val Lys Thr Thr Val Asp Phe Ala Pro Asn
        275                 280                 285

Asn Thr Ile Ile Asn Trp Tyr Val Gly Gly Leu Asn Tyr Gln Thr Val
    290                 295                 300

His His Leu Phe Pro His Ile Cys His Ile His Tyr Pro Lys Ile Ala
305                 310                 315                 320

Pro Ile Leu Ala Glu Val Cys Glu Glu Phe Gly Val Asn Tyr Ala Val
                325                 330                 335

His Gln Thr Phe Phe Gly Ala Leu Ala Ala Asn Tyr Ser Trp Leu Lys
            340                 345                 350

Lys Met Ser Ile Asn Pro Glu Thr Lys Ala Ile Glu Gln
        355                 360                 365
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 21
(D) OTHER INFORMATION: /number= 1
/note= "N=Inosine or Cytosine"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 27
(D) OTHER INFORMATION: /number= 2
/note= "N=Inosine or Cytosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CUACUACUAC UACAYCAYAC NTAYACNAAY AT 32

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic oligonucleotide"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 13
(D) OTHER INFORMATION: /number= 1
/note= "N=Inosine or Cytosine"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 19
(D) OTHER INFORMATION: /number= 2
/note= "N=Inosine or Cytosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CAUCAUCAUC AUNGGRAANA RRTGRTG 27

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
CCAAGCTTCT GCAGGAGCTC TTTTTTTTTT TTTTT                          35
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
His Xaa Xaa His His
1               5
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Gln Xaa Xaa His His
1               5
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 746 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
CGTATGTCAC TCCATTCCAA ACTCGTTCAT GGTATCATAA ATATCAACAC ATTTACGCTC    60
CACTCCTCTA TGGTATTTAC ACACTCAAAT ATCGTACTCA AGATTGGGAA GCTTTTGTAA   120
AGGATGGTAA AAATGGTGCA ATTCGTGTTA GTGTCGCCAC AAATTTCGAT AAGGCCGCTT   180
ACGTCATTGG TAAATTGTCT TTTGTTTTCT TCCGTTTCAT CCTTCCACTC CGTTATCATA   240
GCTTTACAGA TTTAATTTGT TATTTCCTCA TTGCTGAATT CGTCTTTGGT TGGTATCTCA   300
CAATTAATTT CCAAGTTAGT CATGTCGCTG AAGATCTCAA ATTCTTTGCT ACCCCTGAAA   360
GACCAGATGA ACCATCTCAA ATCAATGAAG ATTGGGCAAT CCTTCAACTT AAAACTACTC   420
```

```
AAGATTATGG TCATGGTTCA CTCCTTTGTA CCTTTTTTAG TGGTTCTTTA AATCATCAAG   480
TTGTTCATCA TTTATTCCCA TCAATTGCTC AAGATTTCTA CCCACAACTT GTACCAATTG   540
TAAAAGAAGT TTGTAAAGAA CATAACATTA CTTACCACAT TAAACCAAAC TTCACTGAAG   600
CTATTATGTC ACACATTAAT TACCTTTACA AAATGGGTAA TGATCCAGAT TATGTTAAAA   660
AACCATTAGC CTCAAAAGAT GATTAAATGA AATAACTTAA AAACCAATTA TTTACTTTTG   720
ACAAACAGTA ATATTAATAA ATACAA                                         746
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 227 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Tyr Val Thr Pro Phe Gln Thr Arg Ser Trp Tyr His Lys Tyr Gln
1               5                   10                  15
His Ile Tyr Ala Pro Leu Leu Tyr Gly Ile Tyr Thr Leu Lys Tyr
                20                  25                  30
Arg Thr Gln Asp Trp Glu Ala Phe Val Lys Asp Gly Lys Asn Gly
                35                  40                  45
Ala Ile Arg Val Ser Val Ala Thr Asn Phe Asp Lys Ala Ala Tyr
                50                  55                  60
Val Ile Gly Lys Leu Ser Phe Val Phe Phe Arg Phe Ile Leu Pro
                65                  70                  75
Leu Arg Tyr His Ser Phe Thr Asp Leu Ile Cys Tyr Phe Leu Ile
                80                  85                  90
Ala Glu Phe Val Phe Gly Trp Tyr Leu Thr Ile Asn Phe Gln Val
                95                  100                 105
Ser His Val Ala Glu Asp Leu Lys Phe Phe Ala Thr Pro Glu Arg
                110                 115                 120
Pro Asp Glu Pro Ser Gln Ile Asn Glu Asp Trp Ala Ile Leu Gln
                125                 130                 135
Leu Lys Thr Thr Gln Asp Tyr Gly His Gly Ser Leu Leu Cys Thr
                140                 145                 150
Phe Phe Ser Gly Ser Leu Asn His Gln Val Val His His Leu Phe
                155                 160                 165
Pro Ser Ile Ala Gln Asp Phe Tyr Pro Gln Leu Val Pro Ile Val
                170                 175                 180
Lys Glu Val Cys Lys Glu His Asn Ile Thr Tyr His Ile Lys Pro
                185                 190                 195
Asn Phe Thr Glu Ala Ile Met Ser His Ile Asn Tyr Leu Tyr Lys
                200                 205                 210
Met Gly Asn Asp Pro Asp Tyr Val Lys Lys Pro Leu Ala Ser Lys
                215                 220                 225
Asp Asp ***
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 494 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
TTTTGGAAGG NTCCAAGTTN ACCACGGANT NGGCAAGTTN ACGGGGCGGA AANCGGTTTT    60
CCCCCCAAGC CTTTTGTCGA CTGGTTCTGT GGTGGCTTCC AGTACCAAGT CGACCACCAC   120
TTATTCCCCA GCCTGCCCCG ACACAATCTG GCCAAGACAC ACGCACTGGT CGAATCGTTC   180
TGCAAGGAGT GGGGTGTCCA GTACCACGAA GCCGACCTCG TGGACGGGAC CATGGAAGTC   240
TTGCACCATT TGGGCAGCGT GGCCGGCGAA TTCGTCGTGG ATTTTGTACG CGACGGACCC   300
GCCATGTAAT CGTCGTTCGT GACGATGCAA GGGTTCACGC ACATCTACAC ACACTCACTC   360
ACACAACTAG TGTAACTCGT ATAGAATTCG GTGTCGACCT GGACCTTGTT TGACTGGTTG   420
GGGATAGGGT AGGTAGGCGG ACGCGTGGGT CGNCCCCGGG AATTCTGTGA CCGGTACCTG   480
GCCCGCGTNA AAGT                                                     494
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 87 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Phe Trp Lys Xxx Pro Ser Xxx Pro Arg Xxx Xxx Gln Val Xxx Gly
1               5                   10                  15
Ala Glu Xxx Gly Phe Pro Pro Lys Pro Phe Val Asp Trp Phe Cys
                20                  25                  30
Gly Gly Phe Gln Tyr Gln Val Asp His His Leu Phe Pro Ser Leu
                35                  40                  45
Pro Arg His Asn Leu Ala Lys Thr His Ala Leu Val Glu Ser Phe
                50                  55                  60
Cys Lys Glu Trp Gly Val Gln Tyr His Glu Ala Asp Leu Val Asp
                65                  70                  75
Gly Thr Met Glu Val Leu His His Leu Gly Ser Val Ala Gly Glu
                65                  70                  75
Phe Val Val Asp Phe Val Arg Asp Gly Pro Ala Met
                80                  85
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 520 nucleic acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
GGATGGAGTT CGTCTGGATC GCTGTGCGCT ACGCGACGTG GTTTAAGCGT CATGGGTGCG    60
CTTGGGTACA CGCCGGGGCA GTCGTTGGGC ATGTACTTGT GCGCCTTTGG TCTCGGCTGC   120
ATTTACATTT TTCTGCAGTT CGCCGTAAGT CACACCCATT TGCCCGTGAG CAACCCGGAG   180
GATCAGCTGC ATTGGCTCGA GTACGCGCGG ACCACACTGT GAACATCAGC ACCAAGTCGT   240
```

```
GGTTTGTCAC ATGGTGGATG TCGAACCTCA ACTTTCAGAT CGAGCACCAC CTTTTCCCCA    300
CGGCGCCCCA GTTCCGTTTC AAGGAGATCA GCCCGCGCGT CGAGGCCCTC TTCAAGCGCC    360
ACGGTCTCCC TTACTACGAC ATGCCCTACA CGAGCGCCGT CTCCACCACC TTTGCCAACC    420
TCTACTCCGT CGGCCATTCC GTCGGCGACG CCAAGCGCGA CTAGCCTCTT TTCCTAGACC    480
TTAATTCCCC ACCCCACCCC ATGTTCTGTC TTCCTCCCGC                          520
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 153 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
Met Glu Phe Val Trp Ile Ala Val Arg Tyr Ala Thr Trp Phe Lys
1               5                   10                  15
Arg His Gly Cys Ala Trp Val His Ala Gly Ala Val Val Gly His
                20                  25                  30
Val Leu Val Arg Leu Trp Ser Arg Leu His Leu His Phe Ser Ala
                35                  40                  45
Val Arg Arg Lys Ser His Pro Phe Ala Arg Glu Gln Pro Gly Gly
                50                  55                  60
Ser Ala Ala Leu Ala Arg Val Arg Ala Asp His Thr Val Asn Ile
                65                  70                  75
Ser Thr Lys Ser Trp Phe Val Thr Trp Trp Met Ser Asn Leu Asn
                80                  85                  90
Phe Gln Ile Glu His His Leu Phe Pro Thr Ala Pro Gln Phe Arg
                95                  100                 105
Phe Lys Glu Ile Ser Pro Arg Val Glu Ala Leu Phe Lys Arg His
                110                 115                 120
Gly Leu Pro Tyr Tyr Asp Met Pro Tyr Thr Ser Ala Val Ser Thr
                125                 130                 135
Thr Phe Ala Asn Leu Tyr Ser Val Gly His Ser Val Gly Asp Ala
                140                 145                 150
Lys Arg Asp
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 429 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
ACGCGTCCGC CCACGCGTCC GCCGCGAGCA ACTCATCAAG GAAGGCTACT TTGACCCCTC     60
GCTCCCGCAC ATGACGTACC GCGTGGTCGA GATTGTTGTT CTCTTCGTGC TTTCCTTTTG    120
GCTGATGGGT CAGTCTTCAC CCCTCGCGCT CGCTCTCGGC ATTGTCGTCA GCGGCATCTC    180
TCAGGGTCGC TGCGGCTGGG TAATGCATGA GATGGGCCAT GGGTCGTTCA CTGGTGTCAT    240
TTGGCTTGAC GACCGGTTGT GCGAGTTCTT TTACGGCGTT GGTTGTGGCA TGAGCGGTCA    300
```

TTACTGGAAA AACCAGCACA GCAAACACCA CGCAGCGCCA AACCGGCTCG AGCACGATGT 360
AGATCTCAAC ACCTTGCCAT TGGTGGCCTT CAACGAGCGC GTCGTGCGCA AGGTCCGACC 420

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 125 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
Arg Val Arg Pro Arg Val Arg Arg Glu Gln Leu Ile Lys Glu Gly
1               5                   10                  15
Tyr Phe Asp Pro Ser Leu Pro His Met Thr Tyr Arg Val Val Glu
                20                  25                  30
Ile Val Val Leu Phe Val Leu Ser Phe Trp Leu Met Gly Gln Ser
                35                  40                  45
Ser Pro Leu Ala Leu Ala Leu Gly Ile Val Val Ser Gly Ile Ser
                50                  55                  60
Gln Gly Arg Cys Gly Trp Val Met His Glu Met Gly His Gly Ser
                65                  70                  75
Phe Thr Gly Val Ile Trp Leu Asp Asp Arg Leu Cys Glu Phe Phe
                65                  70                  75
Tyr Gly Val Gly Cys Gly Met Ser Gly His Tyr Trp Lys Asn Gln
                80                  85                  90
His Ser Lys His His Ala Ala Pro Asn Arg Leu Glu His Asp Val
                95                  100                 105
Asp Leu Asn Thr Leu Pro Leu Val Ala Phe Asn Glu Arg Val Val
                110                 115                 120
Arg Lys Val Arg Pro
                125
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1219 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Edited Contig 2692004)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GCACGCCGAC CGGCGCCGGG AGATCCTGGC AAAGTATCCA GAGATAAAGT CCTTGATGAA 60

ACCTGATCCC AATTTGATAT GGATTATAAT TATGATGGTT CTCACCCAGT TGGGTGCATT 120

TTACATAGTA AAAGACTTGG ACTGGAAATG GGTCATATTT GGGGCCTATG CGTTTGGCAG 180

TTGCATTAAC CACTCAATGA CTCTGGCTAT TCATGAGATT GCCCACAATG CTGCCTTTGG 240

CAACTGCAAA GCAATGTGGA ATCGCTGGTT TGGAATGTTT GCTAATCTTC CTATTGGGAT 300

TCCATATTCA ATTTCCTTTA AGAGGTATCA CATGGATCAT CATCGGTACC TTGGAGCTGA 360
TGGCGTCGAT GTAGATATTC CTACCGATTT TGAGGGCTGG TTCTTCTGTA CCGCTTTCAG 420
AAAGTTTATA TGGGTTATTC TTCAGCCTCT CTTTTATGCC TTTCGACCTC TGTTCATCAA 480
CCCCAAACCA ATTACGTATC TGGAAGTTAT CAATACCGTG GCACAGGTCA CTTTTGACAT 540
TTTAATTTAT TACTTTTTGG GAATTAAATC CTTAGTCTAC ATGTTGGCAG CATCTTTACT 600
TGGCCTGGGT TTGCACCCAA TTTCTGGACA TTTTATAGCT GAGCATTACA TGTTCTTAAA 660
GGGTCATGAA ACTTACTCAT ATTATGGGCC TCTGAATTTA CTTACCTTCA ATGTGGGTTA 720
TCATAATGAA CATCATGATT TCCCCAACAT TCCTGGAAAA AGTCTTCCAC TGGTGAGGAA 780
AATAGCAGCT GAATACTATG ACAACCTCCC TCACTACAAT TCCTGGATAA AAGTACTGTA 840
TGATTTTGTG ATGGATGATA CAATAAGTCC CTACTCAAGA ATGAAGAGGC ACCAAAAAGG 900
AGAGATGGTG CTGGAGTAAA TATCATTAGT GCCAAAGGGA TTCTTCTCCA AAACTTTAGA 960
TGATAAAATG GAATTTTTGC ATTATTAAAC TTGAGACCAG TGATGCTCAG AAGCTCCCCT 1020
GGCACAATTT CAGAGTAAGA GCTCGGTGAT ACCAAGAAGT GAATCTGGCT TTTAAACAGT 1080
CAGCCTGACT CTGTACTGCT CAGTTTCACT CACAGGAAAC TTGTGACTTG TGTATTATCG 1140
TCATTGAGGA TGTTTCACTC ATGTCTGTCA TTTTATAAGC ATATCATTTA AAAAGCTTCT 1200
AAAAAGCTAT TTCGCCAGG 1219

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 655 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Edited Contig 2153526)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TTACCTTCTA CGTCCGCTTC TTCCTCACTT ATGTGCCACT ATTGGGGCTG AAAGCTTCCT 60
GGGCCTTTTC TTCATAGTCA GGTTCCTGGA AAGCAACTGG TTTGTGTGGG TGACACAGAT 120
GAACCATATT CCCATGCACA TTGATCATGA CCGGAACATG GACTGGGTTT CCACCCAGCT 180
CCAGGCCACA TGCAATGTCC ACAAGTCTGC CTTCAATGAC TGGTTCAGTG GACACCTCAA 240
CTTCCAGATT GAGCACCATC TTTTTCCCAC GATGCCTCGA CACAATTACC ACAAAGTGGC 300
TCCCCTGGTG CAGTCCTTGT GTGCCAAGCA TGGCATAGAG TACCAGTCCA AGCCCCTGCT 360
GTCAGCCTTC GCCGACATCA TCCACTCACT AAAGGAGTCA GGGCAGCTCT GGCTAGATGC 420
CTATCTTCAC CAATAACAAC AGCCACCCTG CCCAGTCTGG AAGAAGAGGA GGAAGACTCT 480
GGAGCCAAGG CAGAGGGGAG CTTGAGGGAC AATGCCACTA TAGTTTAATA CTCAGAGGGG 540

GTTGGGTTTG GGGACATAAA GCCTCTGACT CAAACTCCTC CCTTTTATCT TCTAGCCACA 600

GTTCTAAGAC CCAAAGTGGG GGGTGGACAC AGAAGTCCCT AGGAGGGAAG GAGCT 655

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 304 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Edited Contig 3506132)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTCTTTTACT TTGGCAATGG CTGGATTCCT ACCCTCATCA CGGCCTTTGT CCTTGCTACC 60

TCTCAGGCCC AAGCTGGATG GCTGCAACAT GATTATGGCC ACCTGTCTGT CTACAGAAAA 120

CCCAAGTGGA ACCACCTTGT CCACAAATTC GTCATTGGCC ACTTAAAGGG TGCCTCTGCC 180

AACTGGTGGA ATCATCGCCA CTTCCAGCAC CACGCCAAGC CTAACATCTT CCACAAGGAT 240

CCCGATGTGA ACATGCTGCA CGTGTTTGTT CTGGGCGAAT GGCAGCCCAT CGAGTACGGC 300

AAGA 304

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 918 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Edited Contig 3854933)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CAGGGACCTA CCCCGCGCTA CTTCACCTGG GACGAGGTGG CCCAGCGCTC AGGGTGCGAG 60

GAGCGGTGGC TAGTGATCGA CCGTAAGGTG TACAACATCA GCGAGTTCAC CCGCCGGCAT 120

CCAGGGGGCT CCCGGGTCAT CAGCCACTAC GCCGGGCAGG ATGCCACGGA TCCCTTTGTG 180

GCCTTCCACA TCAACAAGGG CCTTGTGAAG AAGTATATGA ACTCTCTCCT GATTGGAGAA 240

CTGTCTCCAG AGCAGCCCAG CTTTGAGCCC ACCAAGAATA AAGAGCTGAC AGATGAGTTC 300

CGGGAGCTGC GGGCCACAGT GGAGCGGATG GGGCTCATGA AGGCCAACCA TGTCTTCTTC 360

CTGCTGTACC TGCTGCACAT CTTGCTGCTG GATGGTGCAG CCTGGCTCAC CCTTTGGGTC 420

TTTGGGACGT CCTTTTTGCC CTTCCTCCTC TGTGCGGTGC TGCTCAGTGC AGTTCAGGCC 480

CAGGCTGGCT GGCTGCAGCA TGACTTTGGG CACCTGTCGG TCTTCAGCAC CTCAAAGTGG 540

AACCATCTGC TACATCATTT TGTGATTGGC CACCTGAAGG GGGCCCCCGC CAGTTGGTGG 600

AACCACATGC ACTTCCAGCA CCATGCCAAG CCCAACTGCT TCCGCAAAGA CCCAGACATC 660

AACATGCATC CCTTCTTCTT TGCCTTGGGG AAGATCCTCT CTGTGGAGCT TGGGAAACAG 720

AAGAAAAAAT ATATGCCGTA CAACCACCAG CACARATACT TCTTCCTAAT TGGGCCCCCA 780

GCCTTGCTGC CTCTCTACTT CCAGTGGTAT ATTTTCTATT TTGTTATCCA GCGAAAGAAG 840

TGGGTGGACT TGGCCTGGAT CAGCAAACAG GAATACGATG AAGCCGGGCT TCCATTGTCC 900

ACCGCAAATG CTTCTAAA 918

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1686 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Edited Contig 2511785)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCCACTTAAA GGGTGCCTCT GCCAACTGGT GGAATCATCG CCACTTCCAG CACCACGCCA 60

AGCCTAACAT CTTCCACAAG GATCCCGATG TGAACATGCT GCACGTGTTT GTTCTGGGCG 120

AATGGCAGCC CATCGAGTAC GGCAAGAAGA AGCTGAAATA CCTGCCCTAC AATCACCAGC 180

ACGAATACTT CTTCCTGATT GGGCCGCCGC TGCTCATCCC CATGTATTTC CAGTACCAGA 240

TCATCATGAC CATGATCGTC CATAAGAACT GGGTGGACCT GGCCTGGGCC GTCAGCTACT 300

ACATCCGGTT CTTCATCACC TACATCCCTT TCTACGGCAT CCTGGGAGCC CTCCTTTTCC 360

TCAACTTCAT CAGGTTCCTG GAGAGCCACT GGTTTGTGTG GGTCACACAG ATGAATCACA 420

TCGTCATGGA GATTGACCAG GAGGCCTACC GTGACTGGTT CAGTAGCCAG CTGACAGCCA 480

CCTGCAACGT GGAGCAGTCC TTCTTCAACG ACTGGTTCAG TGGACACCTT AACTTCCAGA 540

TTGAGCACCA CCTCTTCCCC ACCATGCCCC GGCACAACTT ACACAAGATC GCCCCGCTGG 600

TGAAGTCTCT ATGTGCCAAG CATGGCATTG AATACCAGGA GAAGCCGCTA CTGAGGGCCC 660

TGCTGGACAT CATCAGGTCC CTGAAGAAGT CTGGGAAGCT GTGGCTGGAC GCCTACCTTC 720

ACAAATGAAG CCACAGCCCC CGGGACACCG TGGGGAAGGG GTGCAGGTGG GGTGATGGCC 780

AGAGGAATGA TGGGCTTTTG TTCTGAGGGG TGTCCGAGAG GCTGGTGTAT GCACTGCTCA 840

CGGACCCCAT GTTGGATCTT TCTCCCTTTC TCCTCTCCTT TTTCTCTTCA CATCTCCCCC 900

ATAGCACCCT GCCCTCATGG GACCTGCCCT CCCTCAGCCG TCAGCCATCA GCCATGGCCC 960

TCCCAGTGCC TCCTAGCCCC TTCTTCCAAG GAGCAGAGAG GTGGCCACCG GGGGTGGCTC 1020

TGTCCTACCT CCACTCTCTG CCCCTAAAGA TGGGAGGAGA CCAGCGGTCC ATGGGTCTGG 1080

CCTGTGAGTC TCCCCTTGCA GCCTGGTCAC TAGGCATCAC CCCCGCTTTG GTTCTTCAGA 1140

TGCTCTTGGG GTTCATAGGG GCAGGTCCTA GTCGGGCAGG GCCCCTGACC CTCCCGGCCT 1200

GGCTTCACTC TCCCTGACGG CTGCCATTGG TCCACCCTTT CATAGAGAGG CCTGCTTTGT 1260

TACAAAGCTC GGGTCTCCCT CCTGCAGCTC GGTTAAGTAC CCGAGGCCTC TCTTAAGATG 1320
TCCAGGGCCC CAGGCCCGCG GGCACAGCCA GCCCAAACCT TGGGCCCTGG AAGAGTCCTC 1380
CACCCCATCA CTAGAGTGCT CTGACCCTGG GCTTTCACGG GCCCCATTCC ACCGCCTCCC 1440
CAACTTGAGC CTGTGACCTT GGGACCAAAG GGGGAGTCCC TCGTCTCTTG TGACTCAGCA 1500
GAGGCAGTGG CCACGTTCAG GGAGGGGCCG GCTGGCCTGG AGGCTCAGCC CACCCTCCAG 1560
CTTTTCCTCA GGGTGTCCTG AGGTCCAAGA TTCTGGAGCA ATCTGACCCT TCTCCAAAGG 1620
CTCTGTTATC AGCTGGGCAG TGCCAGCCAA TCCCTGGCCA TTTGGCCCCA GGGGACGTGG 1680
GCCCTG 1686

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1843 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Contig 2535)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GTCTTTTACT TTGGCAATGG CTGGATTCCT ACCCTCATCA CGGCCTTTGT CCTTGCTACC 60
TCTCAGGCCC AAGCTGGATG GCTGCAACAT GATTATGGCC ACCTGTCTGT CTACAGAAAA 120
CCCAAGTGGA ACCACCTTGT CCACAAATTC GTCATTGGCC ACTTAAAGGG TGCCTCTGCC 180
AACTGGTGGA ATCATCGCCA CTTCCAGCAC CACGCCAAGC CTAACATCTT CCACAAGGAT 240
CCCGATGTGA ACATGCTGCA CGTGTTTGTT CTGGGCGAAT GGCAGCCCAT CGAGTACGGC 300
AAGAAGAAGC TGAAATACCT GCCCTACAAT CACCAGCACG AATACTTCTT CCTGATTGGG 360
CCGCCGCTGC TCATCCCCAT GTATTTCCAG TACCAGATCA TCATGACCAT GATCGTCCAT 420
AAGAACTGGG TGGACCTGGC CTGGGCCGTC AGCTACTACA TCCGGTTCTT CATCACCTAC 480
ATCCCTTTCT ACGGCATCCT GGGAGCCCTC CTTTTCCTCA ACTTCATCAG GTTCCTGGAG 540
AGCCACTGGT TTGTGTGGGT CACACAGATG AATCACATCG TCATGGAGAT TGACCAGGAG 600
GCCTACCGTG ACTGGTTCAG TAGCCAGCTG ACAGCCACCT GCAACGTGGA GCAGTCCTTC 660
TTCAACGACT GGTTCAGTGG ACACCTTAAC TTCCAGATTG AGCACCACCT CTTCCCCACC 720
ATGCCCCGGC ACAACTTACA CAAGATCGCC CCGCTGGTGA AGTCTCTATG TGCCAAGCAT 780
GGCATTGAAT ACCAGGAGAA GCCGCTACTG AGGGCCCTGC TGGACATCAT CAGGTCCCTG 840
AAGAAGTCTG GGAAGCTGTG GCTGGACGCC TACCTTCACA AATGAAGCCA CAGCCCCCGG 900
GACACCGTGG GGAAGGGGTG CAGGTGGGGT GATGGCCAGA GGAATGATGG GCTTTTGTTC 960
TGAGGGGTGT CCGAGAGGCT GGTGTATGCA CTGCTCACGG ACCCCATGTT GGATCTTTCT 1020

```
CCCTTTCTCC TCTCCTTTTT CTCTTCACAT CTCCCCCATA GCACCCTGCC CTCATGGGAC  1080
CTGCCCTCCC TCAGCCGTCA GCCATCAGCC ATGGCCCTCC CAGTGCCTCC TAGCCCCTTC  1140
TTCCAAGGAG CAGAGAGGTG GCCACCGGGG GTGGCTCTGT CCTACCTCCA CTCTCTGCCC  1200
CTAAAGATGG GAGGAGACCA GCGGTCCATG GGTCTGGCCT GTGAGTCTCC CCTTGCAGCC  1260
TGGTCACTAG GCATCACCCC CGCTTTGGTT CTTCAGATGC TCTTGGGGTT CATAGGGGCA  1320
GGTCCTAGTC GGGCAGGGCC CCTGACCCTC CCGGCCTGGC TTCACTCTCC CTGACGGCTG  1380
CCATTGGTCC ACCCTTTCAT AGAGAGGCCT GCTTTGTTAC AAAGCTCGGG TCTCCCTCCT  1440
GCAGCTCGGT TAAGTACCCG AGGCCTCTCT TAAGATGTCC AGGGCCCCAG GCCCGCGGGC  1500
ACAGCCAGCC CAAACCTTGG GCCCTGGAAG AGTCCTCCAC CCCATCACTA GAGTGCTCTG  1560
ACCCTGGGCT TTCACGGGCC CCATTCCACC GCCTCCCCAA CTTGAGCCTG TGACCTTGGG  1620
ACCAAAGGGG GAGTCCCTCG TCTCTTGTGA CTCAGCAGAG GCAGTGGCCA CGTTCAGGGA  1680
GGGGCCGGCT GGCCTGGAGG CTCAGCCCAC CCTCCAGCTT TTCCTCAGGG TGTCCTGAGG  1740
TCCAAGATTC TGGAGCAATC TGACCCTTCT CCAAAGGCTC TGTTATCAGC TGGGCAGTGC  1800
CAGCCAATCC CTGGCCATTT GGCCCCAGGG GACGTGGGCC CTG                    1843
```

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2257 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid (Edited Contig 253538a)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
CAGGGACCTA CCCCGCGCTA CTTCACCTGG GACGAGGTGG CCCAGCGCTC AGGGTGCGAG    60
GAGCGGTGGC TAGTGATCGA CCGTAAGGTG TACAACATCA GCGAGTTCAC CCGCCGGCAT   120
CCAGGGGGCT CCCGGGTCAT CAGCCACTAC GCCGGGCAGG ATGCCACGGA TCCCTTTGTG   180
GCCTTCCACA TCAACAAGGG CCTTGTGAAG AAGTATATGA ACTCTCTCCT GATTGGAGAA   240
CTGTCTCCAG AGCAGCCCAG CTTTGAGCCC ACCAAGAATA AAGAGCTGAC AGATGAGTTC   300
CGGGAGCTGC GGGCCACAGT GGAGCGGATG GGGCTCATGA AGGCCAACCA TGTCTTCTTC   360
CTGCTGTACC TGCTGCACAT CTTGCTGCTG GATGGTGCAG CCTGGCTCAC CCTTTGGGTC   420
TTTGGGACGT CCTTTTTGCC CTTCCTCCTC TGTGCGGTGC TGCTCAGTGC AGTTCAGCAG   480
GCCCAAGCTG GATGGCTGCA ACATGATTAT GGCCACCTGT CTGTCTACAG AAAACCCAAG   540
TGGAACCACC TTGTCCACAA ATTCGTCATT GGCCACTTAA AGGGTGCCTC TGCCAACTGG   600
TGGAATCATC GCCACTTCCA GCACCACGCC AAGCCTAACA TCTTCCACAA GGATCCCGAT   660
GTGAACATGC TGCACGTGTT TGTTCTGGGC GAATGGCAGC CCATCGAGTA CGGCAAGAAG   720
```

```
AAGCTGAAAT ACCTGCCCTA CAATCACCAG CACGAATACT TCTTCCTGAT TGGGCCGCCG    780
CTGCTCATCC CCATGTATTT CCAGTACCAG ATCATCATGA CCATGATCGT CCATAAGAAC    840
TGGGTGGACC TGGCCTGGGC CGTCAGCTAC TACATCCGGT TCTTCATCAC CTACATCCCT    900
TTCTACGGCA TCCTGGGAGC CCTCCTTTTC CTCAACTTCA TCAGGTTCCT GGAGAGCCAC    960
TGGTTTGTGT GGGTCACACA GATGAATCAC ATCGTCATGG AGATTGACCA GGAGGCCTAC   1020
CGTGACTGGT TCAGTAGCCA GCTGACAGCC ACCTGCAACG TGGAGCAGTC CTTCTTCAAC   1080
GACTGGTTCA GTGGACACCT TAACTTCCAG ATTGAGCACC ACCTCTTCCC CACCATGCCC   1140
CGGCACAACT TACACAAGAT CGCCCCGCTG GTGAAGTCTC TATGTGCCAA GCATGGCATT   1200
GAATACCAGG AGAAGCCGCT ACTGAGGGCC CTGCTGGACA TCATCAGGTC CCTGAAGAAG   1260
TCTGGGAAGC TGTGGCTGGA CGCCTACCTT CACAAATGAA GCCACAGCCC CCGGGACACC   1320
GTGGGGAAGG GGTGCAGGTG GGGTGATGGC CAGAGGAATG ATGGGCTTTT GTTCTGAGGG   1380
GTGTCCGAGA GGCTGGTGTA TGCACTGCTC ACGGACCCCA TGTTGGATCT TTCTCCCTTT   1440
CTCCTCTCCT TTTTCTCTTC ACATCTCCCC CATAGCACCC TGCCCTCATG GGACCTGCCC   1500
TCCCTCAGCC GTCAGCCATC AGCCATGGCC CTCCCAGTGC CTCCTAGCCC CTTCTTCCAA   1560
GGAGCAGAGA GGTGGCCACC GGGGGTGGCT CTGTCCTACC TCCACTCTCT GCCCCTAAAG   1620
ATGGGAGGAG ACCAGCGGTC CATGGGTCTG GCCTGTGAGT CTCCCCTTGC AGCCTGGTCA   1680
CTAGGCATCA CCCCCGCTTT GGTTCTTCAG ATGCTCTTGG GGTTCATAGG GGCAGGTCCT   1740
AGTCGGGCAG GGCCCCTGAC CCTCCCGGCC TGGCTTCACT CTCCCTGACG GCTGCCATTG   1800
GTCCACCCTT TCATAGAGAG GCCTGCTTTG TTACAAAGCT CGGGTCTCCC TCCTGCAGCT   1860
CGGTTAAGTA CCCGAGGCCT CTCTTAAGAT GTCCAGGGCC CCAGGCCCGC GGGCACAGCC   1920
AGCCCAAACC TTGGGCCCTG GAAGAGTCCT CCACCCCATC ACTAGAGTGC TCTGACCCTG   1980
GGCTTTCACG GGCCCCATTC CACCGCCTCC CCAACTTGAG CCTGTGACCT TGGGACCAAA   2040
GGGGGAGTCC CTCGTCTCTT GTGACTCAGC AGAGGCAGTG GCCACGTTCA GGGAGGGGCC   2100
GGCTGGCCTG GAGGCTCAGC CCACCCTCCA GCTTTTCCTC AGGGTGTCCT GAGGTCCAAG   2160
ATTCTGGAGC AATCTGACCC TTCTCCAAAG GCTCTGTTAT CAGCTGGGCA GTGCCAGCCA   2220
ATCCCTGGCC ATTTGGCCCC AGGGGACGTG GGCCCTG                            2257
```

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 411 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 2692004)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

His Ala Asp Arg Arg Arg Glu Ile Leu Ala Lys Tyr Pro Glu Ile
1 5 10 15
Lys Ser Leu Met Lys Pro Asp Pro Asn Leu Ile Trp Ile Ile Ile
20 25 30
Met Met Val Leu Thr Gln Leu Gly Ala Phe Tyr Ile Val Lys Asp
35 40 45
Leu Asp Trp Lys Trp Val Ile Phe Gly Ala Tyr Ala Phe Gly Ser
50 55 60
Cys Ile Asn His Ser Met Thr Leu Ala Ile His Glu Ile Ala His
65 70 75
Asn Ala Ala Phe Gly Asn Cys Lys Ala Met Trp Asn Arg Trp Phe
80 85 90
Gly Met Phe Ala Asn Leu Pro Ile Gly Ile Pro Tyr Ser Ile Ser
95 100 105
Phe Lys Arg Tyr His Met Asp His His Arg Tyr Leu Gly Ala Asp
110 115 120
Gly Val Asp Val Asp Ile Pro Thr Asp Phe Glu Gly Trp Phe Phe
125 130 135
Cys Thr Ala Phe Arg Lys Phe Ile Trp Val Ile Leu Gln Pro Leu
140 145 150
Phe Tyr Ala Phe Arg Pro Leu Phe Ile Asn Pro Lys Pro Ile Thr
155 160 165
Tyr Leu Glu Val Ile Asn Thr Val Ala Gln Val Thr Phe Asp Ile
170 175 180
Leu Ile Tyr Tyr Phe Leu Gly Ile Lys Ser Leu Val Tyr Met Leu
185 190 195
Ala Ala Ser Leu Leu Gly Leu Gly Leu His Pro Ile Ser Gly His
200 205 210
Phe Ile Ala Glu His Tyr Met Phe Leu Lys Gly His Glu Thr Tyr
215 220 225
Ser Tyr Tyr Gly Pro Leu Asn Leu Leu Thr Phe Asn Val Gly Tyr
230 235 240
His Asn Glu His His Asp Phe Pro Asn Ile Pro Gly Lys Ser Leu
245 250 255
Pro Leu Val Arg Lys Ile Ala Ala Glu Tyr Tyr Asp Asn Leu Pro
260 265 270
His Tyr Asn Ser Trp Ile Lys Val Leu Tyr Asp Phe Val Met Asp
275 280 285
Asp Thr Ile Ser Pro Tyr Ser Arg Met Lys Arg His Gln Lys Gly
290 295 300
Glu Met Val Leu Glu *** Ile Ser Leu Val Pro Lys Gly Phe Phe
305 310 315
Ser Lys Thr Leu Asp Asp Lys Met Glu Phe Leu His Tyr *** Thr
320 325 330
*** Asp Gln *** Cys Ser Glu Ala Pro Leu Ala Gln Phe Gln Ser
335 340 345
Lys Ser Ser Val Ile Pro Arg Ser Glu Ser Gly Phe *** Thr Val
350 355 360
Ser Leu Thr Leu Tyr Cys Ser Val Ser Leu Thr Gly Asn Leu ***
365 370 375
Leu Val Tyr Tyr Arg His *** Gly Cys Phe Thr His Val Cys His
380 385 390
Phe Ile Ser Ile Ser Phe Lys Lys Leu Leu Lys Ser Tyr Phe Ala
400 405 410
Arg

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 218 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 2153526)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```
Tyr Leu Leu Arg Pro Leu Leu Pro His Leu Cys Ala Thr Ile Gly
1               5                   10                  15
Ala Glu Ser Phe Leu Gly Leu Phe Phe Ile Val Arg Phe Leu Glu
                20                  25                  30
Ser Asn Trp Phe Val Trp Val Thr Gln Met Asn His Ile Pro Met
                35                  40                  45
His Ile Asp His Asp Arg Asn Met Asp Trp Val Ser Thr Gln Leu
                50                  55                  60
Gln Ala Thr Cys Asn Val His Lys Ser Ala Phe Asn Asp Trp Phe
                65                  70                  75
Ser Gly His Leu Asn Phe Gln Ile Glu His His Leu Phe Pro Thr
                80                  85                  90
Met Pro Arg His Asn Tyr His Lys Val Ala Pro Leu Val Gln Ser
                95                  100                 105
Leu Cys Ala Lys His Gly Ile Glu Tyr Gln Ser Lys Pro Leu Leu
                110                 115                 120
Ser Ala Phe Ala Asp Ile Ile His Ser Leu Lys Glu Ser Gly Gln
                125                 130                 135
Leu Trp Leu Asp Ala Tyr Leu His Gln *** Gln Gln Pro Pro Cys
                140                 145                 150
Pro Val Trp Lys Lys Arg Arg Lys Thr Leu Glu Pro Arg Gln Arg
                155                 160                 165
Gly Ala *** Gly Thr Met Pro Leu *** Phe Asn Thr Gln Arg Gly
                170                 175                 180
Leu Gly Leu Gly Thr *** Ser Leu *** Leu Lys Leu Leu Pro Phe
                185                 190                 195
Ile Phe *** Pro Gln Phe *** Asp Pro Lys Trp Gly Val Asp Thr
                200                 205                 210
Glu Val Pro Arg Arg Glu Gly Ala
                215
```

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 71 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 3506132)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
Val Phe Tyr Phe Gly Asn Gly Trp Ile Pro Thr Leu Ile Thr Ala
1               5                   10                  15
Phe Val Leu Ala Thr Ser Gln Ala Gln Ala Gly Trp Leu Gln His
                20                  25                  30
Asp Tyr Gly His Leu Ser Val Tyr Arg Lys Pro Lys Trp Asn His
                35                  40                  45
Leu Val His Lys Phe Val Ile Gly His Leu Lys Gly Ala Ser Ala
```

```
                        50                            55                            60
Asn Trp Trp Asn His Arg His Phe Gln His His Ala Lys Pro Asn
                        65                            70                            75
Leu Gly Glu Trp Gln Pro Ile Glu Tyr Gly Lys Xxx
                        80                            85
```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 306 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 3854933)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```
Gln Gly Pro Thr Pro Arg Tyr Phe Thr Trp Asp Glu Val Ala Gln
1               5                   10                  15
Arg Ser Gly Cys Glu Glu Arg Trp Leu Val Ile Asp Arg Lys Val
                20                  25                  30
Tyr Asn Ile Ser Glu Phe Thr Arg Arg His Pro Gly Gly Ser Arg
                35                  40                  45
Val Ile Ser His Tyr Ala Gly Gln Asp Ala Thr Asp Pro Phe Val
                50                  55                  60
Ala Phe His Ile Asn Lys Gly Leu Val Lys Lys Tyr Met Asn Ser
                65                  70                  75
Leu Leu Ile Gly Glu Leu Ser Pro Glu Gln Pro Ser Phe Glu Pro
                80                  85                  90
Thr Lys Asn Lys Glu Leu Thr Asp Glu Phe Arg Glu Leu Arg Ala
                95                  100                 105
Thr Val Glu Arg Met Gly Leu Met Lys Ala Asn His Val Phe Phe
                110                 115                 120
Leu Leu Tyr Leu Leu His Ile Leu Leu Leu Asp Gly Ala Ala Trp
                125                 130                 135
Leu Thr Leu Trp Val Phe Gly Thr Ser Phe Leu Pro Phe Leu Leu
                140                 145                 150
Cys Ala Val Leu Leu Ser Ala Val Gln Ala Gln Ala Gly Trp Leu
                155                 160                 165
Gln His Asp Phe Gly His Leu Ser Val Phe Ser Thr Ser Lys Trp
                170                 175                 180
Asn His Leu Leu His His Phe Val Ile Gly His Leu Lys Gly Ala
                185                 190                 195
Pro Ala Ser Trp Trp Asn His Met His Phe Gln His His Ala Lys
                200                 205                 210
Pro Asn Cys Phe Arg Lys Asp Pro Asp Ile Asn Met His Pro Phe
                215                 220                 225
Phe Phe Ala Leu Gly Lys Ile Leu Ser Val Glu Leu Gly Lys Gln
                230                 235                 240
Lys Lys Lys Tyr Met Pro Tyr Asn His Gln His Xxx Tyr Phe Phe
                245                 250                 255
Leu Ile Gly Pro Pro Ala Leu Leu Pro Leu Tyr Phe Gln Trp Tyr
                260                 265                 270
Ile Phe Tyr Phe Val Ile Gln Arg Lys Lys Trp Val Asp Leu Ala
                275                 280                 285
Trp Ile Ser Lys Gln Glu Tyr Asp Glu Ala Gly Leu Pro Leu Ser
                290                 295                 300
Thr Ala Asn Ala Ser Lys
```

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 566 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 2511785)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
His Leu Lys Gly Ala Ser Ala Asn Trp Trp Asn His Arg His Phe
1               5                   10                  15
Gln His His Ala Lys Pro Asn Ile Phe His Lys Asp Pro Asp Val
                20                  25                  30
Asn Met Leu His Val Phe Val Leu Gly Glu Trp Gln Pro Ile Glu
                35                  40                  45
Tyr Gly Lys Lys Lys Leu Lys Tyr Leu Pro Tyr Asn His Gln His
                50                  55                  60
Glu Tyr Phe Phe Leu Ile Gly Pro Pro Leu Leu Ile Pro Met Tyr
                65                  70                  75
Phe Gln Tyr Gln Ile Ile Met Thr Met Ile Val His Lys Asn Trp
                80                  85                  90
Val Asp Leu Ala Trp Ala Val Ser Tyr Tyr Ile Arg Phe Phe Ile
                95                  100                 105
Thr Tyr Ile Pro Phe Tyr Gly Ile Leu Gly Ala Leu Leu Phe Leu
                110                 115                 120
Asn Phe Ile Arg Phe Leu Glu Ser His Trp Phe Val Trp Val Thr
                125                 130                 135
Gln Met Asn His Ile Val Met Glu Ile Asp Gln Glu Ala Tyr Arg
                140                 145                 150
Asp Trp Phe Ser Ser Gln Leu Thr Ala Thr Cys Asn Val Glu Gln
                155                 160                 165
Ser Phe Phe Asn Asp Trp Phe Ser Gly His Leu Asn Phe Gln Ile
                170                 175                 180
Glu His His Leu Phe Pro Thr Met Pro Arg His Asn Leu His Lys
                185                 190                 195
Ile Ala Pro Leu Val Lys Ser Leu Cys Ala Lys His Gly Ile Glu
                200                 205                 210
Tyr Gln Glu Lys Pro Leu Leu Arg Ala Leu Leu Asp Ile Ile Arg
                215                 220                 225
Ser Leu Lys Lys Ser Gly Lys Leu Trp Leu Asp Ala Tyr Leu His
                230                 235                 240
Lys *** Ser His Ser Pro Arg Asp Thr Val Gly Lys Gly Cys Arg
                245                 250                 255
Trp Gly Asp Gly Gln Arg Asn Asp Gly Leu Leu Phe *** Gly Val
                260                 265                 270
Ser Glu Arg Leu Val Tyr Ala Leu Leu Thr Asp Pro Met Leu Asp
                275                 280                 285
Leu Ser Pro Phe Leu Leu Ser Phe Phe Ser Ser His Leu Pro His
                290                 295                 300
Ser Thr Leu Pro Ser Trp Asp Leu Pro Ser Leu Ser Arg Gln Pro
                305                 310                 315
Ser Ala Met Ala Leu Pro Val Pro Pro Ser Pro Phe Phe Gln Gly
                320                 325                 330
Ala Glu Arg Trp Pro Pro Gly Val Ala Leu Ser Tyr Leu His Ser
                335                 340                 345
```

Leu Pro Leu Lys Met Gly Gly Asp Gln Arg Ser Met Gly Leu Ala
350 355 360
Cys Glu Ser Pro Leu Ala Ala Trp Ser Leu Gly Ile Thr Pro Ala
365 370 375
Leu Val Leu Gln Met Leu Leu Gly Phe Ile Gly Ala Gly Pro Ser
380 385 390
Arg Ala Gly Pro Leu Thr Leu Pro Ala Trp Leu His Ser Pro ***
400 405 410
Arg Leu Pro Leu Val His Pro Phe Ile Glu Arg Pro Ala Leu Leu
415 420 425
Gln Ser Ser Gly Leu Pro Pro Ala Ala Arg Leu Ser Thr Arg Gly
430 435 440
Leu Ser *** Asp Val Gln Gly Pro Arg Pro Ala Gly Thr Ala Ser
445 450 455
Pro Asn Leu Gly Pro Trp Lys Ser Pro Pro Pro His His *** Ser
460 465 470
Ala Leu Thr Leu Gly Phe His Gly Pro His Ser Thr Ala Ser Pro
475 480 485
Thr *** Ala Cys Asp Leu Gly Thr Lys Gly Gly Val Pro Arg Leu
490 495 500
Leu *** Leu Ser Arg Gly Ser Gly His Val Gln Gly Gly Ala Gly
505 510 515
Trp Pro Gly Gly Ser Ala His Pro Pro Ala Phe Pro Gln Gly Val
520 525 530
Leu Arg Ser Lys Ile Leu Glu Gln Ser Asp Pro Ser Pro Lys Ala
535 540 545
Leu Leu Ser Ala Gly Gln Cys Gln Pro Ile Pro Gly His Leu Ala
550 555 560
Pro Gly Asp Val Gly Pro Xxx
565

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 619 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 2535)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Val Phe Tyr Phe Gly Asn Gly Trp Ile Pro Thr Leu Ile Thr Ala
1 5 10 15
Phe Val Leu Ala Thr Ser Gln Ala Gln Ala Gly Trp Leu Gln His
20 25 30
Asp Tyr Gly His Leu Ser Val Tyr Arg Lys Pro Lys Trp Asn His
35 40 45
Leu Val His Lys Phe Val Ile Gly His Leu Lys Gly Ala Ser Ala
50 55 60
Asn Trp Trp Asn His Arg His Phe Gln His His Ala Lys Pro Asn
65 70 75
Ile Phe His Lys Asp Pro Asp Val Asn Met Leu His Val Phe Val
80 85 90
Leu Gly Glu Trp Gln Pro Ile Glu Tyr Gly Lys Lys Lys Leu Lys
95 100 105
Tyr Leu Pro Tyr Asn His Gln His Glu Tyr Phe Phe Leu Ile Gly
110 115 120

```
Pro Pro Leu Leu Ile Pro Met Tyr Phe Gln Tyr Gln Ile Ile Met
                125                 130                 135
Thr Met Ile Val His Lys Asn Trp Val Asp Leu Ala Trp Ala Val
                140                 145                 150
Ser Tyr Tyr Ile Arg Phe Phe Ile Thr Tyr Ile Pro Phe Tyr Gly
                155                 160                 165
Ile Leu Gly Ala Leu Leu Phe Leu Asn Phe Ile Arg Phe Leu Glu
                170                 175                 180
Ser His Trp Phe Val Trp Val Thr Gln Met Asn His Ile Val Met
                185                 190                 195
Glu Ile Asp Gln Glu Ala Tyr Arg Asp Trp Phe Ser Ser Gln Leu
                200                 205                 210
Thr Ala Thr Cys Asn Val Glu Gln Ser Phe Phe Asn Asp Trp Phe
                215                 220                 225
Ser Gly His Leu Asn Phe Gln Ile Glu His His Leu Phe Pro Thr
                230                 235                 240
Met Pro Arg His Asn Leu His Lys Ile Ala Pro Leu Val Lys Ser
                245                 250                 255
Leu Cys Ala Lys His Gly Ile Glu Tyr Gln Glu Lys Pro Leu Leu
                260                 265                 270
Arg Ala Leu Leu Asp Ile Ile Arg Ser Leu Lys Lys Ser Gly Lys
                275                 280                 285
Leu Trp Leu Asp Ala Tyr Leu His Lys *** Ser His Ser Pro Arg
                290                 295                 300
Asp Thr Val Gly Lys Gly Cys Arg Trp Gly Asp Gly Gln Arg Asn
                305                 310                 315
Asp Gly Leu Leu Phe *** Gly Val Ser Glu Arg Leu Val Tyr Ala
                320                 325                 330
Leu Leu Thr Asp Pro Met Leu Asp Leu Ser Pro Phe Leu Leu Ser
                335                 340                 345
Phe Phe Ser Ser His Leu Pro His Ser Thr Leu Pro Ser Trp Asp
                350                 355                 360
Leu Pro Ser Leu Ser Arg Gln Pro Ser Ala Met Ala Leu Pro Val
                365                 370                 375
Pro Pro Ser Pro Phe Phe Gln Gly Ala Glu Arg Trp Pro Pro Gly
                380                 385                 390
Val Ala Leu Ser Tyr Leu His Ser Leu Pro Leu Lys Met Gly Gly
                400                 405                 410
Asp Gln Arg Ser Met Gly Leu Ala Cys Glu Ser Pro Leu Ala Ala
                415                 420                 425
Trp Ser Leu Gly Ile Thr Pro Ala Leu Val Leu Gln Met Leu Leu
                430                 435                 440
Gly Phe Ile Gly Ala Gly Pro Ser Arg Ala Gly Pro Leu Thr Leu
                445                 450                 455
Pro Ala Trp Leu His Ser Pro *** Arg Leu Pro Leu Val His Pro
                460                 465                 470
Phe Ile Glu Arg Pro Ala Leu Leu Gln Ser Ser Gly Leu Pro Pro
                475                 480                 485
Ala Ala Arg Leu Ser Thr Arg Gly Leu Ser *** Asp Val Gln Gly
                490                 495                 500
Pro Arg Pro Ala Gly Thr Ala Ser Pro Asn Leu Gly Pro Trp Lys
                505                 510                 515
Ser Pro Pro Pro His His *** Ser Ala Leu Thr Leu Gly Phe His
                520                 525                 530
Gly Pro His Ser Thr Ala Ser Pro Thr *** Ala Cys Asp Leu Gly
                535                 540                 545
Thr Lys Gly Gly Val Pro Arg Leu Leu *** Leu Ser Arg Gly Ser
                550                 555                 560
Gly His Val Gln Gly Gly Ala Gly Trp Pro Gly Gly Ser Ala His
                565                 570                 575
Pro Pro Ala Phe Pro Gln Gly Val Leu Arg Ser Lys Ile Leu Glu
                580                 585                 590
```

```
Gln Ser Asp Pro Ser Pro Lys Ala Leu Leu Ser Ala Gly Gln Cys
                595                 600                 605
Gln Pro Ile Pro Gly His Leu Ala Pro Gly Asp Val Gly Pro Xxx
                610                 615                 620
```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 757 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid (Translation of Contig 253538a)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
Gln Gly Pro Thr Pro Arg Tyr Phe Thr Trp Asp Glu Val Ala Gln
1               5                   10                  15
Arg Ser Gly Cys Glu Glu Arg Trp Leu Val Ile Asp Arg Lys Val
                20                  25                  30
Tyr Asn Ile Ser Glu Phe Thr Arg Arg His Pro Gly Gly Ser Arg
                35                  40                  45
Val Ile Ser His Tyr Ala Gly Gln Asp Ala Thr Asp Pro Phe Val
                50                  55                  60
Ala Phe His Ile Asn Lys Gly Leu Val Lys Lys Tyr Met Asn Ser
                65                  70                  75
Leu Leu Ile Gly Glu Leu Ser Pro Glu Gln Pro Ser Phe Glu Pro
                80                  85                  90
Thr Lys Asn Lys Glu Leu Thr Asp Glu Phe Arg Glu Leu Arg Ala
                95                  100                 105
Thr Val Glu Arg Met Gly Leu Met Lys Ala Asn His Val Phe Phe
                110                 115                 120
Leu Leu Tyr Leu Leu His Ile Leu Leu Leu Asp Gly Ala Ala Trp
                125                 130                 135
Leu Thr Leu Trp Val Phe Gly Thr Ser Phe Leu Pro Phe Leu Leu
                140                 145                 150
Cys Ala Val Leu Leu Ser Ala Val Gln Gln Ala Gln Ala Gly Trp
                155                 160                 165
Leu Gln His Asp Tyr Gly His Leu Ser Val Tyr Arg Lys Pro Lys
                170                 175                 180
Trp Asn His Leu Val His Lys Phe Val Ile Gly His Leu Lys Gly
                185                 190                 195
Ala Ser Ala Asn Trp Trp Asn His Arg His Phe Gln His His Ala
                200                 205                 210
Lys Pro Asn Ile Phe His Lys Asp Pro Asp Val Asn Met Leu His
                215                 220                 225
Val Phe Val Leu Gly Glu Trp Gln Pro Ile Glu Tyr Gly Lys Lys
                230                 235                 240
Lys Leu Lys Tyr Leu Pro Tyr Asn His Gln His Glu Tyr Phe Phe
                245                 250                 255
Leu Ile Gly Pro Pro Leu Leu Ile Pro Met Tyr Phe Gln Tyr Gln
                260                 265                 270
Ile Ile Met Thr Met Ile Val His Lys Asn Trp Val Asp Leu Ala
                275                 280                 285
Trp Ala Val Ser Tyr Tyr Ile Arg Phe Phe Ile Thr Tyr Ile Pro
                290                 295                 300
Phe Tyr Gly Ile Leu Gly Ala Leu Leu Phe Leu Asn Phe Ile Arg
                305                 310                 315
```

```
Phe Leu Glu Ser His Trp Phe Val Trp Val Thr Gln Met Asn His
                320                 325                 330
Ile Val Met Glu Ile Asp Gln Glu Ala Tyr Arg Asp Trp Phe Ser
                335                 340                 345
Ser Gln Leu Thr Ala Thr Cys Asn Val Glu Gln Ser Phe Phe Asn
                350                 355                 360
Asp Trp Phe Ser Gly His Leu Asn Phe Gln Ile Glu His His Leu
                365                 370                 375
Phe Pro Thr Met Pro Arg His Asn Leu His Lys Ile Ala Pro Leu
                380                 385                 390
Val Lys Ser Leu Cys Ala Lys His Gly Ile Glu Tyr Gln Glu Lys
                400                 405                 410
Pro Leu Leu Arg Ala Leu Leu Asp Ile Ile Arg Ser Leu Lys Lys
                415                 420                 425
Ser Gly Lys Leu Trp Leu Asp Ala Tyr Leu His Lys *** Ser His
                430                 435                 440
Ser Pro Arg Asp Thr Val Gly Lys Gly Cys Arg Trp Gly Asp Gly
                445                 450                 455
Gln Arg Asn Asp Gly Leu Leu Phe *** Gly Val Ser Glu Arg Leu
                460                 465                 470
Val Tyr Ala Leu Leu Thr Asp Pro Met Leu Asp Leu Ser Pro Phe
                475                 480                 485
Leu Leu Ser Phe Phe Ser Ser His Leu Pro His Ser Thr Leu Pro
                490                 495                 500
Ser Trp Asp Leu Pro Ser Leu Ser Arg Gln Pro Ser Ala Met Ala
                505                 510                 515
Leu Pro Val Pro Pro Ser Pro Phe Phe Gln Gly Ala Glu Arg Trp
                520                 525                 530
Pro Pro Gly Val Ala Leu Ser Tyr Leu His Ser Leu Pro Leu Lys
                535                 540                 545
Met Gly Gly Asp Gln Arg Ser Met Gly Leu Ala Cys Glu Ser Pro
                550                 555                 560
Leu Ala Ala Trp Ser Leu Gly Ile Thr Pro Ala Leu Val Leu Gln
                565                 570                 575
Met Leu Leu Gly Phe Ile Gly Ala Gly Pro Ser Arg Ala Gly Pro
                580                 585                 590
Leu Thr Leu Pro Ala Trp Leu His Ser Pro *** Arg Leu Pro Leu
                595                 600                 605
Val His Pro Phe Ile Glu Arg Pro Ala Leu Leu Gln Ser Ser Gly
                610                 615                 620
Leu Pro Pro Ala Ala Arg Leu Ser Thr Arg Gly Leu Ser *** Asp
                625                 630                 635
Val Gln Gly Pro Arg Pro Ala Gly Thr Ala Ser Pro Asn Leu Gly
                640                 645                 650
Pro Trp Lys Ser Pro Pro Pro His His *** Ser Ala Leu Thr Leu
                655                 660                 665
Gly Phe His Gly Pro His Ser Thr Ala Ser Pro Thr *** Ala Cys
                670                 675                 680
Asp Leu Gly Thr Lys Gly Gly Val Pro Arg Leu Leu *** Leu Ser
                685                 690                 695
Arg Gly Ser Gly His Val Gln Gly Gly Ala Gly Trp Pro Gly Gly
                700                 705                 710
Ser Ala His Pro Pro Ala Phe Pro Gln Gly Val Leu Arg Ser Lys
                715                 720                 725
Ile Leu Glu Gln Ser Asp Pro Ser Pro Lys Ala Leu Leu Ser Ala
                730                 735                 740
Gly Gln Cys Gln Pro Ile Pro Gly His Leu Ala Pro Gly Asp Val
                745                 750                 755
Gly Pro Xxx
```

What is claimed is:

1. An isolated nucleic acid comprising:

a nucleotide sequence depicted in a SEQ ID NO. 1

2. A polypeptide encoded by said nucleic acid of claim 1.

3. A purified or isolated polypeptide comprising an amino acid sequence depicted in SEQ ID NO: 2.

4. An isolated nucleic acid encoding the polypeptide of SEQ ID NO: 2.

5. An isolated nucleic acid comprising:

a nucleotide sequence which encodes a polypeptide that desaturates a fatty acid molecule at carbon 5 from the carboxyl end of said fatty acid molecule.

6. The isoloated nucleic acid according to Claim 5, wherein said nucleotide sequence is derived from eukaryotic cell.

7. The isolated nucleic acid according to Claim 6, wherein said eukaryotic cell is a fungal cell.

8. The isolated nucleic acid according to Claim 7, wherein said fungal cell is of the genus *Mortierella.*

9. The isolated nucleic acid according to Claim 8, wherein said *Mortierella* cell is of the species *Mortierella alpina.*

10. The isolated nucleic acid according to Claim 5, wherein said nucleotide sequence anneals to a nucleotide sequence depicted in SEQ ID NO: 1.

11. The nucleic acid of claim 10, wherein said nucleotide sequence encodes an amino acid sequence depicted in SEQ ID NO: 2.

12. The nucleic acid of claim 11, wherein said amino acid sequence depicted in SEQ ID NO: 2 is selected from the group consisting of amino acid residues 30-38, 41-44, 171-175, 203-212, and 387-394.

13. An isolated or purified polypeptide which desaturates a fatty acid molecule at carbon 5 from the carboxyl end of said fatty acid molecule.

14. An isolated nucleic acid comprising:

a nucleotide sequence which is substantially identical to a sequence of at least 50 nucleotides in SEQ ID NO 1.

15. An isolated nucleic acid sequence having at least about 50% identity to SEQ ID NO 1.

16. A nucleic acid construct comprising:

a nucleotide sequence depicted in a SEQ ID NO: 1 linked to a heterologous nucleic acid.

17. A nucleic acid construct comprising:

a nucleotide sequence depicted in a SEQ ID NO: 1 operably linked to a promoter.

18. The nucleic acid construct of claim 17, wherein said promoter is functional in a microbial cell.

19. The nucleic acid construct of claim 18, wherein said microbial cell is a yeast cell.

20. The nucleic acid construct of claim 17, wherein said nucleotide sequence is derived from a fungus.

21. The nucleic acid according to Claim 19, wherein said fungus is of the genus *Mortierella*.

22. The nucleic acid according to Claim 20, wherein said fungus is of the species *Mortierella alpina*.

23. A nucleic acid construct comprising:

a nucleotide sequence which encodes a polypeptide comprising an amino acid sequence which corresponds to or is complementary to an amino acid sequence depicted in SEQ ID NO: 2, wherein said nucleotide sequence is operably linked to a promoter which is functional in a host cell, wherein said nucleotide

sequence encodes a polypeptide which desaturates a fatty acid molecule at carbon 5 from the carboxyl end of a fatty acid molecule.

24. A nucleic acid construct comprising:

a nucleotide sequence which encodes a functionally active Δ5-desaturase, said desaturase having an amino acid sequence which corresponds to or is complementary to all of or a portion of an amino acid sequence depicted in a SEQ ID NO: 2, wherein said nucleotide sequence is operably linked to a promoter functional in a host cell.

25. A recombinant yeast cell comprising:

a nucleic construct according to Claim 23 or Claim 24.

26. The recombinant yeast cell according to Claim 25, wherein said yeast cell is a *Saccharomyces* cell.

27. A host cell comprising:

at least one copy of a nucleotide sequence which encodes a polypeptide which converts dihomo-γ-linolenic acid to arachidonic acid, wherein said microbial cell or an ancestor of said microbial cell was transformed with a vector comprising said nucleotide sequence, and wherein said nucleotide sequence is operably linked to a promoter functional in said host cell.

28. The microbial cell according to Claim 27, wherein said cell is a host cell selected from the group consisting of a fungal cell and an algal cell.

29. The microbial cell according to Claim 28, wherein said fungal cell is a yeast cell and said algae cell is marine algal cell.

30. The microbial cell according to Claim 27, wherein said cell is enriched for 20:3 fatty acids as compared to a host cell which is devoid of said nucleotide sequence.

31. The microbial cell according to Claim 27, wherein said cell is enriched for 20:4 or ω-3 20:4 fatty acids as compared to a host cell which is devoid of said DNA sequence.

32. The microbial cell according to Claim 27, wherein said cell is enriched for 20:5 fatty acids as compared to a host cell which is devoid of said DNA sequence.

33. The microbial cell according to Claim 27, wherein said cell has an altered amount of 20:3 (8, 11, 14) fatty acid as compared to an untransformed microbial cell.

34. A method for production of arachidonic acid in a microbial cell culture, said method comprising:

growing a microbial cell culture having a plurality of microbial cells, wherein said microbial cells or ancestors of said microbial cells were transformed with a vector comprising one or more nucleic acids having a nucleotide sequence which encodes a polypeptide which converts dihomo-γ-linolenic acid to arachidonic acid, wherein said one or more nucleic acids are operably linked to a promoter, under conditions wherein said one or more nucleic acids are expressed and arachidonic acid is produced in said microbial cell culture.

35. The method of Claim 34, wherein said polypeptide is an enzyme which desaturates a fatty acid molecule at carbon 5 from the carboxyl end of said fatty acid molecule.

36. The method of Claim 34, wherein said nucleotide sequence is derived from a *Mortierella species.*

37. The method according to Claim 34, wherein said dihomo-γ-linolenic acid is exogenously supplied.

38. The method according to Claim 34, wherein said microbial cells are yeast cells.

39. The method according to Claim 38, wherein said yeast cells are *Saccharomyces species* cells.

40. The method according to Claim 34, wherein said conditions are inducible.

41. A recombinant yeast cell which converts greater than about 5% of a 20:3 fatty acid to a 20:4 fatty acid.

42. A nucleic acid probe comprising:

a nucleotide sequence as represented by SEQ ID NO:1.

43. A host cell comprising:

a nucleic acid construct according to Claim 23 or Claim 24.

44. A host cell comprising:

a vector which includes a nucleic acid which encodes a fatty acid desaturase derived from *Mortierella alpina*, wherein said fatty acid desaturase comprises an amino acid sequence represented by SEQ ID NO:2, wherein said nucleic acid is operably linked to a promoter.

45. The host cell according to Claim 44, wherein said host cell is a eukaryotic cell.

46. The host cell according to Claim 45, wherein said eukaryotic cell is selected from the group consisting of a mammalian cell, a plant cell, a fungal cell, an avian cell and an algal cell.

47. The host cell according to Claim 45, wherein said host cell contains dihomo-gamma-linolenic acid.

48. The host cell according to Claim 45, wherein said host cell contains EPA.

49. The host cell according to Claim 44, wherein said promoter is exogenously supplied.

50. A method for desaturating a dihomo-γ-linolenic acid, said method comprising:

culturing a recombinant microbial cell according to Claim 37, under conditions suitable for expression of polypeptide encoded by said nucleic acid, wherein said host cell further comprises a fatty acid substrate of said polypeptide.

51. A fatty acid desaturated by the method according to Claim 50.

52. An oil comprising a fatty acid according to Claim 51.

53. A method for obtaining altered long chain polyunsaturated fatty acid biosynthesis comprising the steps of:

growing a microbe having cells which contain a transgene which encodes a transgene expression product which desaturates a fatty acid molecule at carbon 5 from the carboxyl end of said fatty acid molecule, wherein said trangene is operably associated with an expression control sequence, under conditions whereby said transgene is expressed, whereby long chain polyunsaturated fatty acid biosynthesis in said cells is altered.

54. A method for obtaining altered long chain polyunsaturated fatty acid biosynthesis comprising the steps of:

growing a microbe having cells which contain a transgene, derived from a fungus or algae, which encodes a transgene expression product which desaturates a fatty acid molecule at carbon 5 from the carboxyl end of said fatty acid molecule, wherein said trangene is operably associated with an expression control sequence, under conditions whereby said one or more transgenes is expressed, whereby long chain polyunsaturated fatty acid biosynthesis in said cells is altered.

55. The method according to claims 53 or 54, wherein said long chain polyunsaturated fatty acid is selected from the group consisting of ARA, DGLA and EPA.

56. A microbal oil or fraction thereof produced according to the method of claims 53 or 54.

57. A method of treating or preventing malnutrition comprising administering said microbal oil of claim 56 to a patient in need of said treatment or prevention in an amount sufficient to effect said treatment or prevention.

58. A pharmaceutical composition comprising said microbal oil or fraction of claim 56 and a pharmaceutically acceptable carrier.

59. The pharmaceutical composition of claim 58, wherein said pharmaceutical composition is in the form of a solid or a liquid.

60. The pharmaceutical composition of claim 59, wherein said pharmaceutical composition is in a capsule or tablet form.

61. The pharmaceutical composition of claim 58 further comprising at least one nutrient selected from the group consisting of a vitamin, a mineral, a carbohydrate, a sugar, an amino acid, a free fatty acid, a phospholipid, an antioxidant, and a phenolic compound.

62. A nutritional formula comprising said microbal oil or fraction thereof of claim 56.

63. The nutritional formula of claim 62, wherein said nutritional formula is selected from the group consisting of an infant formula, a dietary supplement, and a dietary substitute.

64. The nutritional formula of claim 63, wherein said infant formula, dietary supplement or dietary supplement is in the form of a liquid or a solid.

65. An infant formula comprising said microbal oil or fraction thereof of claim 56.

66. The infant formula of claim 65 further comprising at least one macronutrient selected from the group consisting of coconut oil, soy oil, canola oil, mono- and diglycerides, glucose, edible lactose, electrodialysed whey, electrodialysed skim milk, milk whey, soy protein, and other protein hydrolysates.

67. The infant formula of claim 66 further comprising at least one vitamin selected from the group consisting of Vitamins A, C, D, E, and B complex; and at least one mineral selected from the group consisting of calcium, magnesium, zinc, manganese, sodium, potassium, phosphorus, copper, chloride, iodine, selenium, and iron.

68. A dietary supplement comprising said microbal oil or fraction thereof of claim 56.

69. The dietary supplement of claim 68 further comprising at least one macronutrient selected from the group consisting of coconut oil, soy oil, canola oil, mono- and diglycerides, glucose, edible lactose, electrodialysed whey, electrodialysed skim milk, milk whey, soy protein, and other protein hydrolysates.

70. The dietary supplement of claim 69 further comprising at least one vitamin selected from the group consisting of Vitamins A, C, D, E, and B complex; and at least one mineral selected from the group consisting of calcium, magnesium, zinc, manganese, sodium, potassium, phosphorus, copper, chloride, iodine, selenium, and iron.

71. The dietary supplement of claim 68 or claim 70, wherein said dietary supplement is administered to a human or an animal.

72. A dietary substitute comprising said microbal oil or fraction thereof of claim 56.

73. The dietary substitute of claim 72 further comprising at least one macronutrient selected from the group consisting of coconut oil, soy oil, canola oil, mono- and diglycerides, glucose, edible lactose, electrodialysed whey, electrodialysed skim milk, milk whey, soy protein, and other protein hydrolysates.

74. The dietary substitute of claim 73 further comprising at least one vitamin selected from the group consisting of Vitamins A, C, D, E, and B complex; and at least one mineral selected from the group consisting of calcium, magnesium, zinc, manganese, sodium, potassium, phosphorus, copper, chloride, iodine, selenium, and iron.

75. The dietary substitute of claim 72 or claim 74, wherein said dietary substitute is administered to a human or animal.

76. A method of treating a patient having a condition caused by insuffient intake or production of polyunsaturated fatty acids comprising administering to said patient said dietary substitute of claim 72 or said dietary supplement of claim 68 in an amount sufficient to effect said treatment.

77. The method of claim 72, wherein said dietary substitute or said dietary supplement is administered enterally or parenterally.

78. A cosmetic comprising said microbal oil or fraction thereof of claim 56.

79. The cosmetic of claim 78, wherein said cosmetic is applied topically.

80. The pharmaceutical composition of claim 58, wherein said pharmaceutical composition is administered to a human or an animal.

81. An animal feed comprising said microbal oil or fraction thereof of claim 56.

82. The method of claim 54 wherein said fungus is *Mortierella species*.

83. The method of claim 82 wherein said fungus is *Mortierella alpina*.

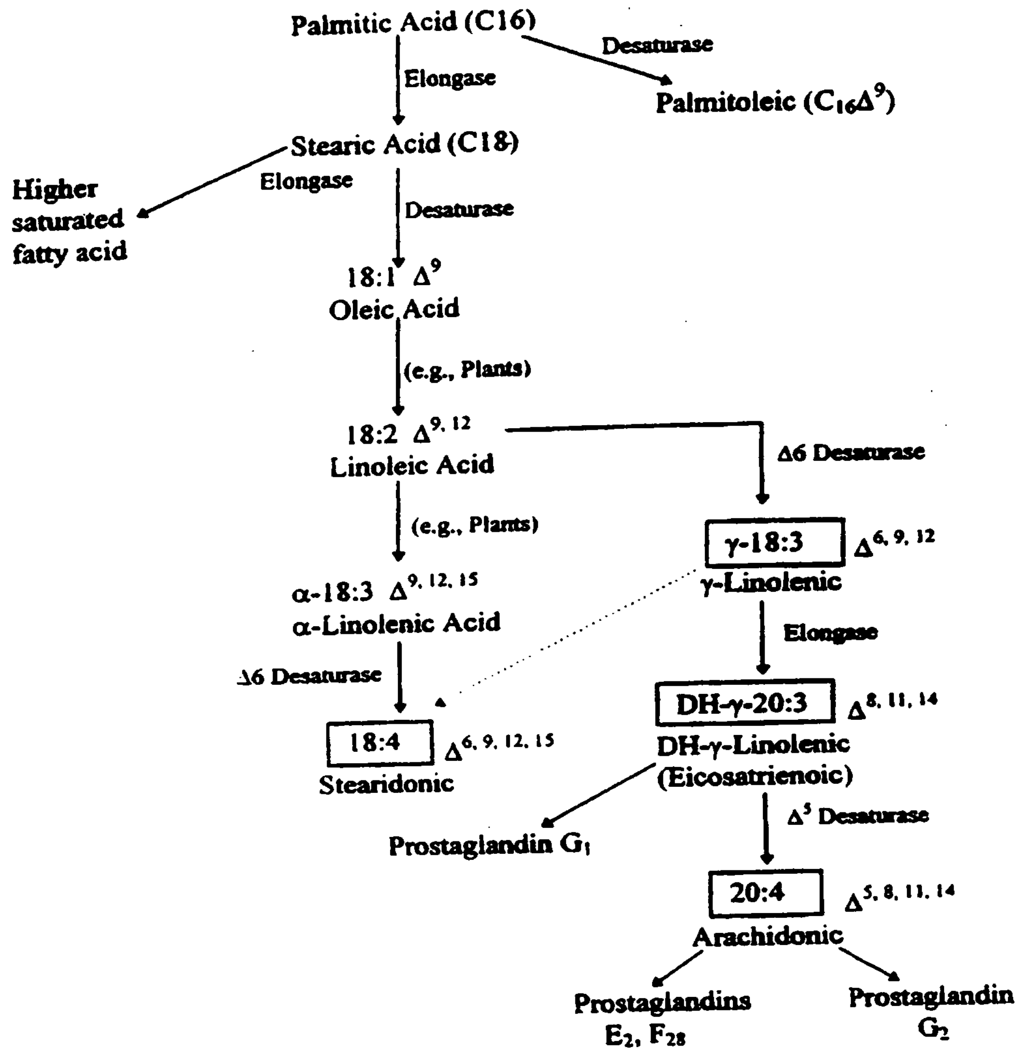
84. An isolated nucleotide sequence selected from the group consisting of SEQ ID NO:13 and SEQ ID NO:15.

85. An isolated nucleotide sequence from the group consisting of SEQ ID NO:7 and SEQ ID NO:19.

86. An isolated nucleotide sequence comprising a nucleotide sequence selected from the group consisting of: SEQ ID NO:13; SEQ ID NO: 15; SEQ ID NO:17; SEQ ID NO:19; SEQ ID NO:21; SEQ ID NO:22; SEQ ID NO:23; SEQ ID NO:24; SEQ ID NO:25; SEQ ID NO:26 and SEQ ID NO:27.

87. An isolated peptide sequence comprising a peptide sequence selected from the group consisting of: SEQ ID NO:14; SEQ ID NO:16; SEQID NO:18; SEQ ID NO:20; SEQ ID NO:28; SEQ ID NO:29; SEQ ID NO:30; SEQ ID NO:31; SEQ ID NO:32; SEQ ID NO:33 and SEQ ID NO:34.

88. Purified polypeptides produced from the nucleotide sequences of claims 84-86.



Palmitic Acid (C16)
Desaturase
Palmitoleic ($C_{16}\Delta^9$)
Elongase
Stearic Acid (C18)
Elongase
Higher saturated fatty acid
Desaturase
18:1 Δ^9
Oleic Acid
(e.g., Plants)
18:2 $\Delta^{9,12}$
Linoleic Acid
Δ6 Desaturase
(e.g., Plants)
γ-18:3 $\Delta^{6,9,12}$
γ-Linolenic
α-18:3 $\Delta^{9,12,15}$
α-Linolenic Acid
Elongase
Δ6 Desaturase
DH-γ-20:3 $\Delta^{8,11,14}$
18:4 $\Delta^{6,9,12,15}$
Stearidonic
DH-γ-Linolenic
(Eicosatrienoic)
Δ^5 Desaturase
Prostaglandin G_1
20:4 $\Delta^{5,8,11,14}$
Arachidonic
Prostaglandins E_2, F_{28}
Prostaglandin G_2

FIG. 1

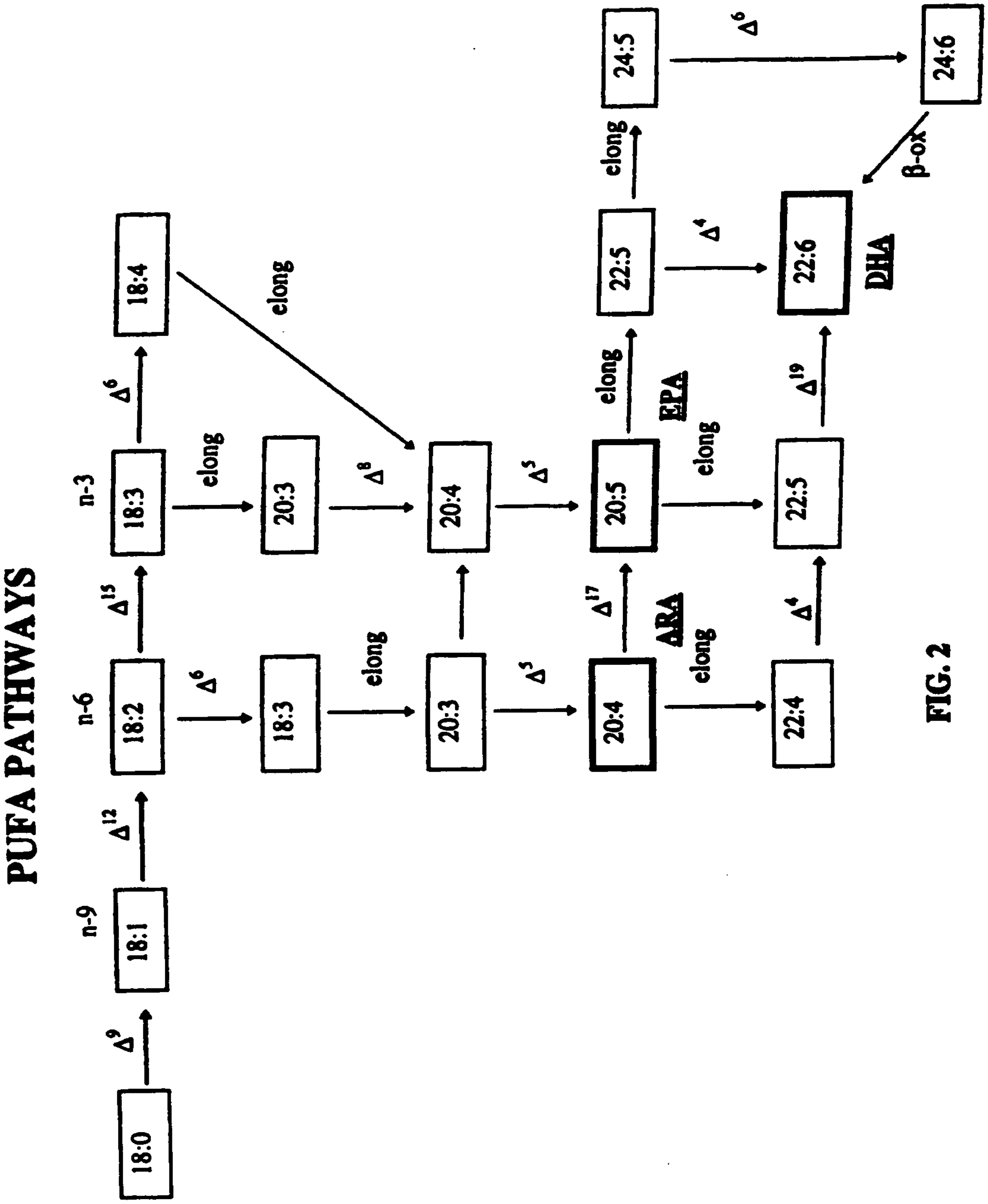

PUFA PATHWAYS
18:0
Δ9
n-9
18:1
Δ12
n-6
18:2
Δ6
18:3
elong
20:3
Δ5
20:4
ARA
elong
22:4
Δ15
Δ17
Δ4
n-3
18:3
elong
20:3
Δ8
20:4
Δ5
20:5
EPA
elong
22:5
Δ19
Δ6
18:4
elong
elong
22:5
Δ4
22:6
DHA
elong
24:5
Δ6
24:6
β-ox

FIG. 2

FIG. 3A

```
GCTTCCTCCA GTTCATCCTC CATTTCGCCA CCTGCATTCT TTACGACCGT TAAGCAAG

60
*
ATG GGA ACG GAC CAA GGA AAA ACC TTC ACC TGG GAA GAG CTG GCG GCC
Met Gly Thr Asp Gln Gly Lys Thr Phe Thr Trp Glu Glu Leu Ala Ala

                120
                 *
CAT AAC ACC AAG GAC GAC CTA CTC TTG GCC ATC CGC GGC AGG GTG TAC
His Asn Thr Lys Asp Asp Leu Leu Leu Ala Ile Arg Gly Arg Val Tyr

                                180
                                 *
GAT GTC ACA AAG TTC TTG AGC CGC CAT CCT GGT GGA GTG GAC ACT CTC
Asp Val Thr Lys Phe Leu Ser Arg His Pro Gly Gly Val Asp Thr Leu

                                                240
                                                 *
CTG CTC GGA GCT GGC CGA GAT GTT ACT CCG GTC TTT GAG ATG TAT CAC
Leu Leu Gly Ala Gly Arg Asp Val Thr Pro Val Phe Glu Met Tyr His

GCG TTT GGG GCT GCA GAT GCC ATT ATG AAG AAG TAC TAT GTC GGT ACA
Ala Phe Gly Ala Ala Asp Ala Ile Met Lys Lys Tyr Tyr Val Gly Thr

300
*
CTG GTC TCG AAT GAG CTG CCC ATC TTC CCG GAG CCA ACG GTG TTC CAC
Leu Val Ser Asn Glu Leu Pro Ile Phe Pro Glu Pro Thr Val Phe His

                360
                 *
AAA ACC ATC AAG ACG AGA GTC GAG GGC TAC TTT ACG GAT CGG AAC ATT
Lys Thr Ile Lys Thr Arg Val Glu Gly Tyr Phe Thr Asp Arg Asn Ile
```

FIG. 3B

```
                                      420
                                        *
GAT CCC AAG AAT AGA CCA GAG ATC TGG GGA CGA TAC GCT CTT ATC TTT
Asp Pro Lys Asn Arg Pro Glu Ile Trp Gly Arg Tyr Ala Leu Ile Phe

                                                      480
                                                        *
GGA TCC TTG ATC GCT TCC TAC TAC GCG CAG CTC TTT GTG CCT TTC GTT
Gly Ser Leu Ile Ala Ser Tyr Tyr Ala Gln Leu Phe Val Pro Phe Val

GTC GAA CGC ACA TGG CTT CAG GTG GTG TTT GCA ATC ATC ATG GGA TTT
Val Glu Arg Thr Trp Leu Gln Val Val Phe Ala Ile Ile Met Gly Phe

540
 *
GCG TGC GCA CAA GTC GGA CTC AAC CCT CTT CAT GAT GCG TCT CAC TTT
Ala Cys Ala Gln Val Gly Leu Asn Pro Leu His Asp Ala Ser His Phe

          600
            *
TCA GTG ACC CAC AAC CCC ACT GTC TGG AAG ATT CTG GGA GCC ACG CAC
Ser Val Thr His Asn Pro Thr Val Trp Lys Ile Leu Gly Ala Thr His

                                      660
                                        *
GAC TTT TTC AAC GGA GCA TCG TAC CTG GTG TGG ATG TAC CAA CAT ATG
Asp Phe Phe Asn Gly Ala Ser Tyr Leu Val Trp Met Tyr Gln His Met

                                                      720
                                                        *
CTC GGC CAT CAC CCC TAC ACC AAC ATT GCT GGA GCA GAT CCC GAC GTG
Leu Gly His His Pro Tyr Thr Asn Ile Ala Gly Ala Asp Pro Asp Val
```

FIG. 3C

```
TCG ACG TCT GAG CCC GAT GTT CGT CGT ATC AAG CCC AAC CAA AAG TGG
Ser Thr Ser Glu Pro Asp Val Arg Arg Ile Lys Pro Asn Gln Lys Trp

780
 *
TTT GTC AAC CAC ATC AAC CAG CAC ATG TTT GTT CCT TTC CTG TAC GGA
Phe Val Asn His Ile Asn Gln His Met Phe Val Pro Phe Leu Tyr Gly

                840
                 *
CTG CTG GCG TTC AAG GTG CGC ATT CAG GAC ATC AAC ATT TTG TAC TTT
Leu Leu Ala Phe Lys Val Arg Ile Gln Asp Ile Asn Ile Leu Tyr Phe

                                900
                                 *
GTC AAG ACC AAT GAC GCT ATT CGT GTC AAT CCC ATC TCG ACA TGG CAC
Val Lys Thr Asn Asp Ala Ile Arg Val Asn Pro Ile Ser Thr Trp His

                                                960
                                                 *
ACT GTG ATG TTC TGG GGC GGC AAG GCT TTC TTT GTC TGG TAT CGC CTG
Thr Val Met Phe Trp Gly Gly Lys Ala Phe Phe Val Trp Tyr Arg Leu

ATT GTT CCC CTG CAG TAT CTG CCC CTG GGC AAG GTG CTG CTC TTG TTC
Ile Val Pro Leu Gln Tyr Leu Pro Leu Gly Lys Val Leu Leu Leu Phe

1020
 *
ACG GTC GCG GAC ATG GTG TCG TCT TAC TGG CTG GCG CTG ACC TTC CAG
Thr Val Ala Asp Met Val Ser Ser Tyr Trp Leu Ala Leu Thr Phe Gln
```

FIG. 3D

```
                 1080
                  •
GCG AAC CAC GTT GTT GAG GAA GTT CAG TGG CCG TTG CCT GAC GAG AAC
Ala Asn His Val Val Glu Glu Val Gln Trp Pro Leu Pro Asp Glu Asn

                             1140
                              •
GGG ATC ATC CAA AAG GAC TGG GCA GCT ATG CAG GTC GAG ACT ACG CAG
Gly Ile Ile Gln Lys Asp Trp Ala Ala Met Gln Val Glu Thr Thr Gln
                                             1200
                                              •
GAT TAC GCA CAC GAT TCG CAC CTC TGG ACC AGC ATC ACT GGC AGC TTG
Asp Tyr Ala His Asp Ser His Leu Trp Thr Ser Ile Thr Gly Ser Leu

AAC TAC CAG GCT GTG CAC CAT CTG TTC CCC AAC GTG TCG CAG CAC CAT
Asn Tyr Gln Ala Val His His Leu Phe Pro Asn Val Ser Gln His His

1260
 •
TAT CCC GAT ATT CTG GCC ATC ATC AAG AAC ACC TGC AGC GAG TAC AAG
Tyr Pro Asp Ile Leu Ala Ile Ile Lys Asn Thr Cys Ser Glu Tyr Lys

             1320
              •
GTT CCA TAC CTT GTC AAG GAT ACG TTT TGG CAA GCA TTT GCT TCA CAT
Val Pro Tyr Leu Val Lys Asp Thr Phe Trp Gln Ala Phe Ala Ser His

                              1380
                               •
TTG GAG CAC TTG CGT GTT CTT GGA CTC CGT CCC AAG GAA GAG TAGA
Leu Glu His Leu Arg Val Leu Gly Leu Arg Pro Lys Glu Glu

                                        1440
                                         •
AGAAAAAAAG CGCCGAATGA AGTATTGCCC CCTTTTTCTC CAAGAATGGC AAAAGGAGAT

CAAGTGGACA TTCTCTATGA AGA
```

FIG. 4

10	20	30	40	50	60
LHHTYTNIAG	ADPDVSTSEP	DVRRIKPNQK	WFVNHINQHM	FVPFLYGLLA	FKVRIQDINI
70	80	90	100	110	120
LYFVKTNDAI	RVNPISTWHT	VMFWGGKAFF	VWYRLIVPLQ	YLPLGKVLLL	FTVADMVSSY
130	140	150	160	170	180
WLALTFQANY	VVEEVQWPLP	DENGIIQKDW	AAMQVETTQD	YAHDSHLWTS	ITGSLNYQXV
HHLFPH					

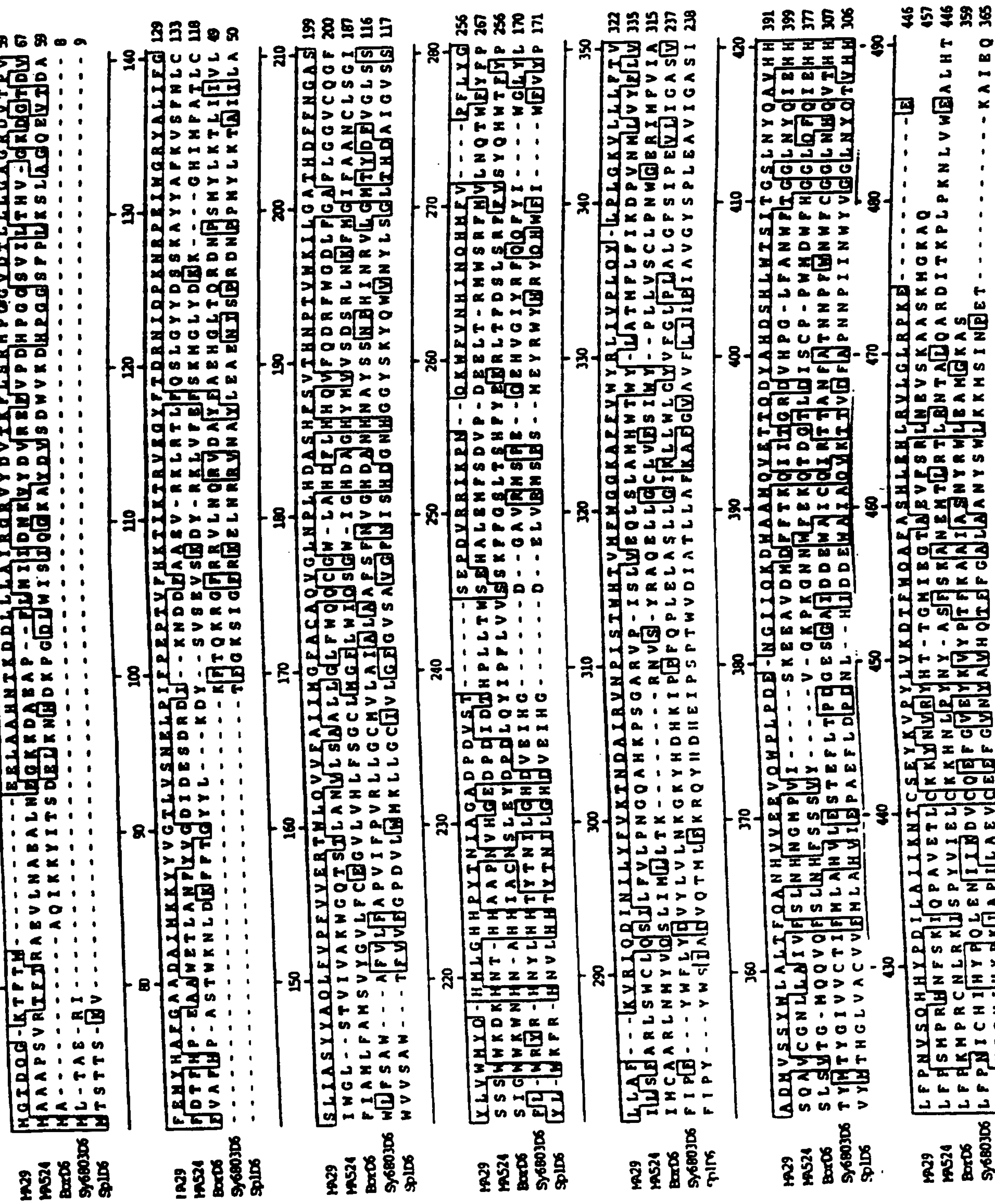


FIG. 5

Effect of Timing of Substrate Addition on Expression of *Mortierella* Delta-5-desaturase Gene in Yeast (SC334)

(Induction Temperature 15 C)

A

B

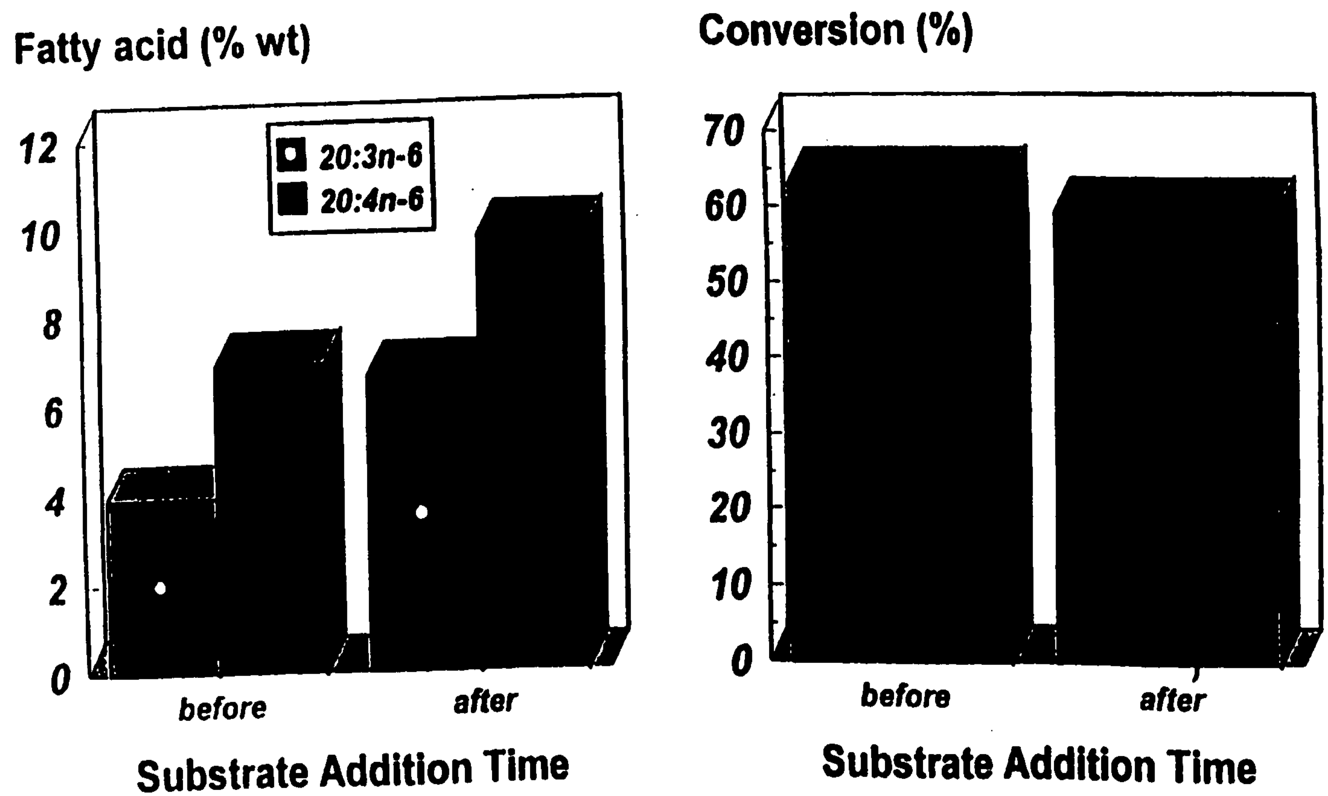


FIG. 6

Effect of Concentration of Inducer (Galactose) on Expression of *Mortierella* Delta-5-desaturase Gene in Yeast (SC334)

(Induction Temperature 15 C)

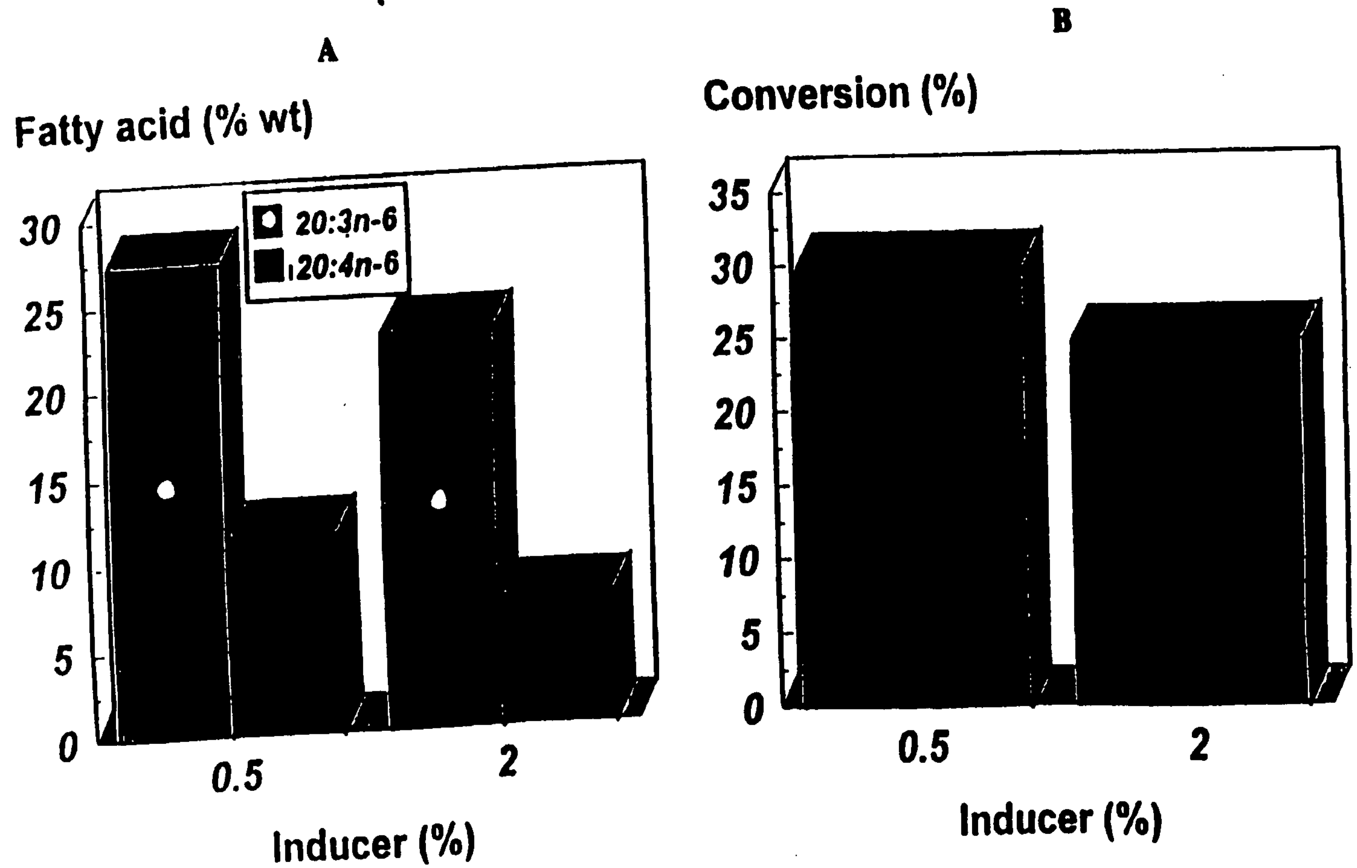


FIG. 7

Effect of Induction Temperature on Expression of *Mortierella* Delta-5-desaturase Gene in Yeast (Strain SC334)

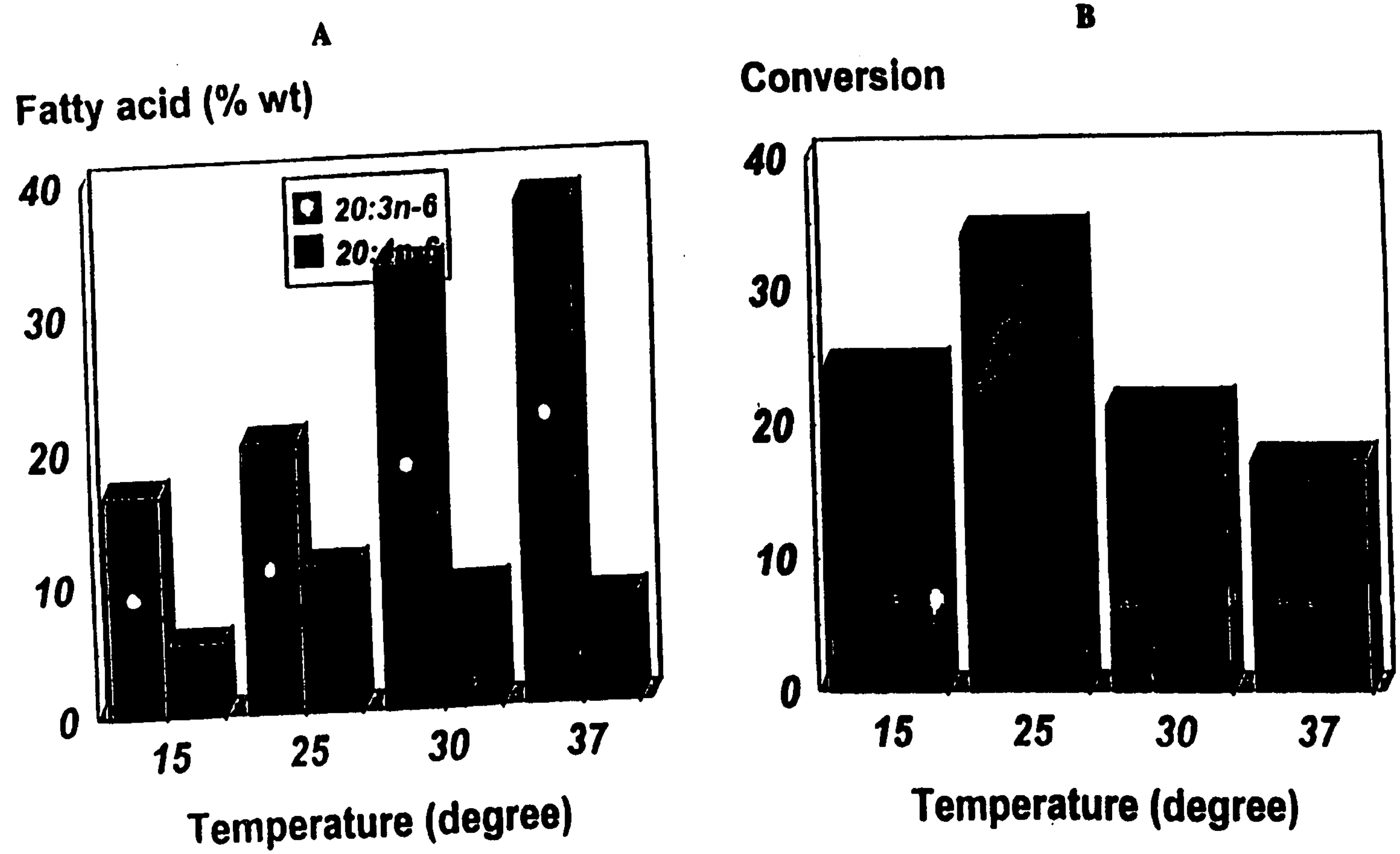


FIG. 8

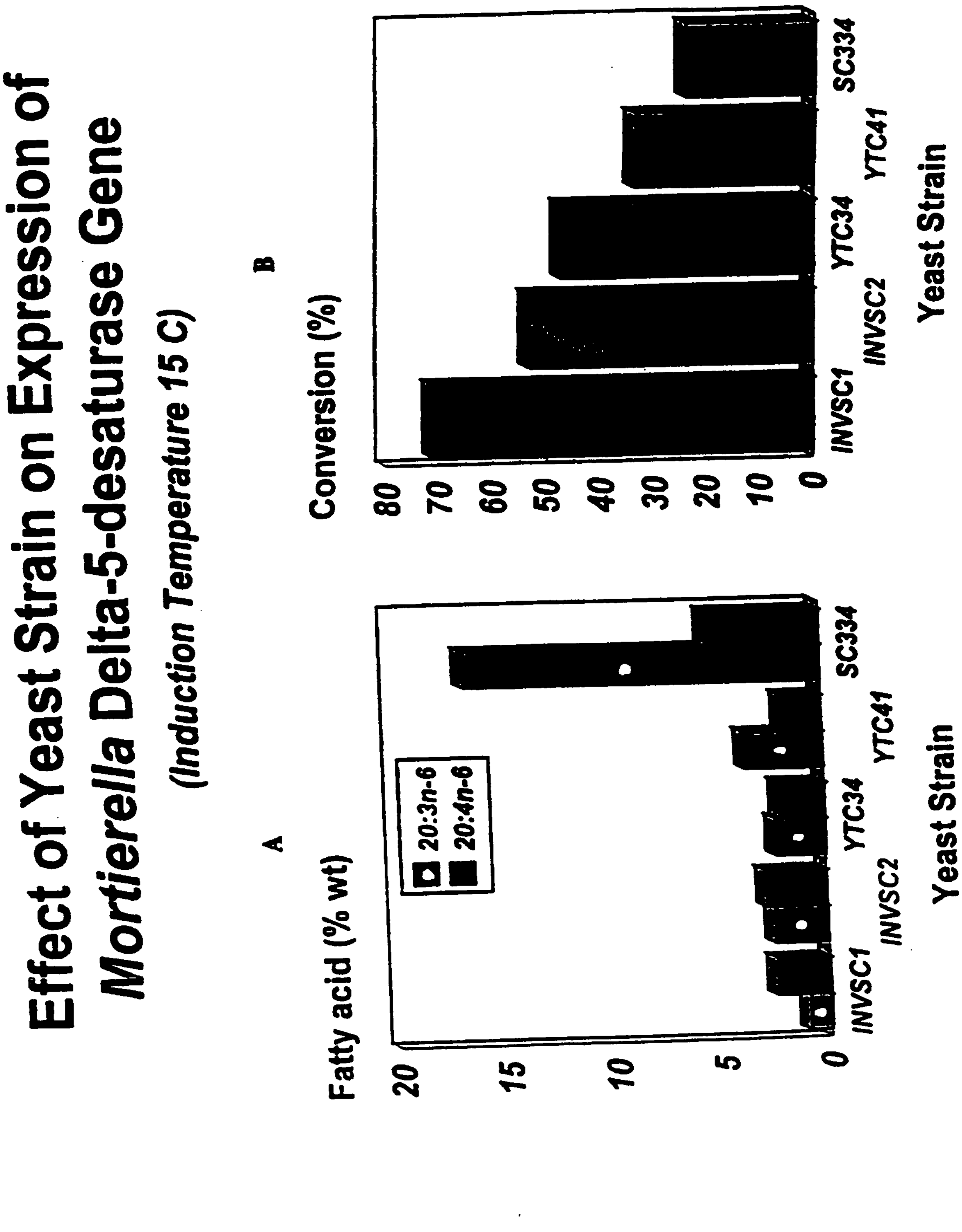

Effect of Yeast Strain on Expression of
Mortierella Delta-5-desaturase Gene
(Induction Temperature 15 C)
A
Fatty acid (% wt)
20:3n-6
20:4n-6
20
15
10
5
0
INVSC1
INVSC2
YTC34
YTC41
SC334
Yeast Strain
B
Conversion (%)
80
70
60
50
40
30
20
10
0
INVSC1
INVSC2
YTC34
YTC41
SC334
Yeast Strain

FIG.9

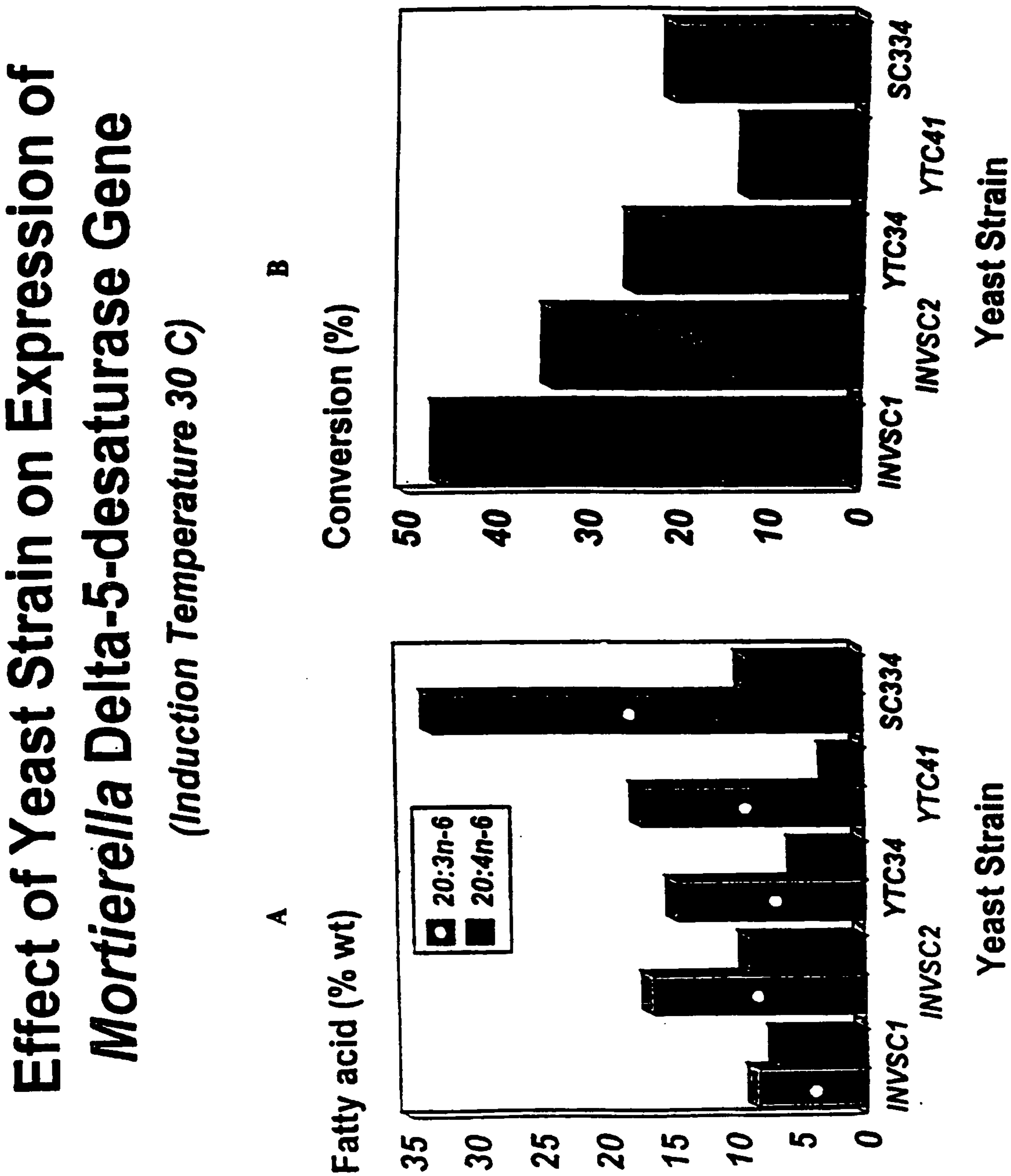

Effect of Yeast Strain on Expression of
Mortierella Delta-5-desaturase Gene
(Induction Temperature 30 C)
A
Fatty acid (% wt)
35
30
25
20
15
10
5
0
20:3n-6
20:4n-6
INVSC1
INVSC2
YTC34
YTC41
SC334
Yeast Strain
B
Conversion (%)
50
40
30
20
10
0
INVSC1
INVSC2
YTC34
YTC41
SC334
Yeast Strain

FIG. 10

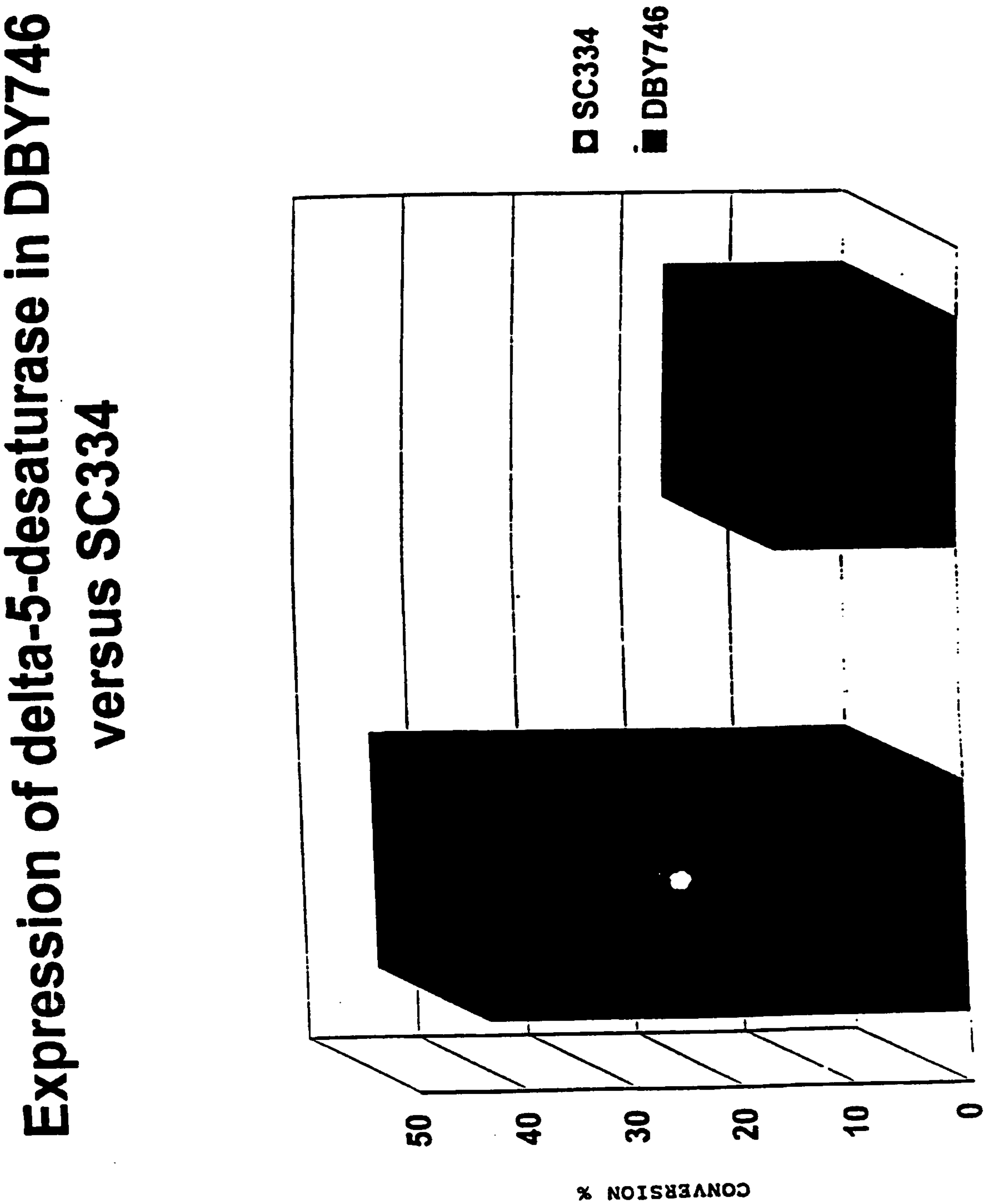

Expression of delta-5-desaturase in DBY746
versus SC334
SC334
DBY746
CONVERSION %
50
40
30
20
10
0

FIG. 11

Effect of Medium Composition and Temperature on Expression of delta-5-desaturase Activity in Two Host Strains (SC334 and INVSC1)

A

SC344

B

INVSC1

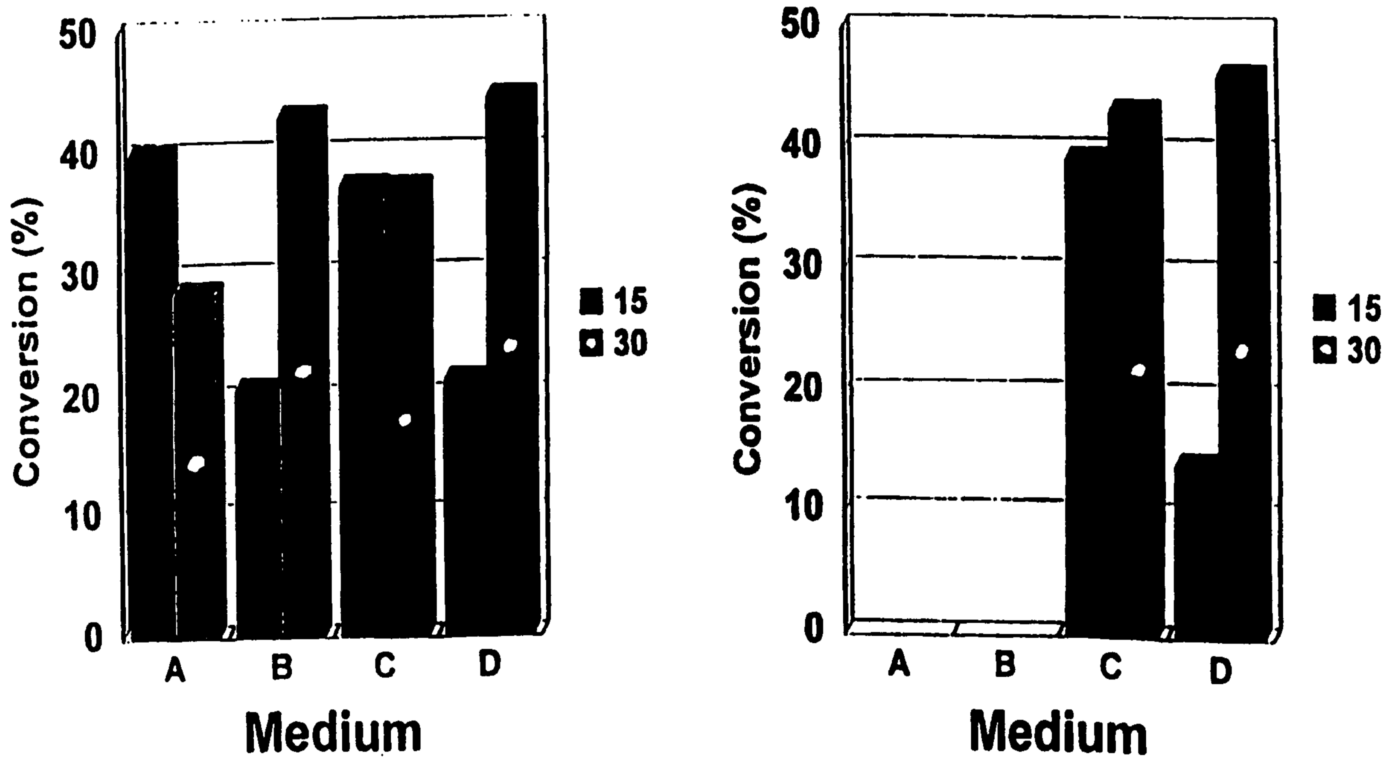


FIG. 12

FastA Match of ma29 and c ntig 253538a

```
SCORES       Init1:    117 Initn:    225 Opt:    256
Smith-Waterman score: 408;     27.0% identity in 441 aa overlap

                       10        20        30        40        50
ma29gcg.pep   MGTDQGKT---FTWEELAAHNTKDDLLLAIRGRVYDVTKFLSRHPGGVDTLLLGAGRDVT
                 || |   |||:|:| ::   ::  :|  :||::::| ||||| ::   ||:|:|
253538a          QGPTPRYFTWDEVAQRSGCEERWLVIDRKVYNISEFTRRHPGGSRVISHYAGQDAT
                       10        20        30        40        50

                60        70        80        90       100       110
ma29gcg.pep   PVFEMYHAF-GAADAIMKKYYVGTLVSNELPIFPEPTVFHKTIKTRVEGYFTDRNIDPKN
               |  :|   | :   |::  :| | | | | | |||  ::        |   | : :   :
253538a       DPFVAFHINKGLVKKYMNSLLIGEL-SPEQPSF-EPTKNKELTDEFRELRATVERMGLMK
                60        70        80         90       100       110

               120       130       140       150        160       170
ma29gcg.pep   RPEIWGRYALIFGSLIASYYAQLFVPFVVERTWLQVVF-AIIMGFACAQVGLNPLHDASH
               :::  : | :  ::      |  :: :|   ::|  :: |:::: : ||:|    || :|
253538a       ANHVF--FLLYLLHILLLDGAAWLTLWVFGTSFLPFLLCAVLLSAVQAQAGWLQ-HDYGH
                120       130       140       150       160       170

                180       190           200       210       220
ma29gcg.pep   FSVTHNPTVWKILGATHDF----FNGASYLVWMYQHMLGHHPYTNIAGADPDVSTSE---
              :|| ::|  |: |  :| |    ::|||  | ::| : ||  ||  ||||:   :
253538a       LSVYRKPK-WNHL--VHKFVIGHLKGASANWWNHRH-FQHHAKPNIFHKDPDVNMLHVFV
                 180        190       200       210       220

                 230       240        250         260       270       280
ma29gcg.pep   ----PDVRRIKPNQKWF-VNHINQHMFV--PFLYGLLAFKVRIQDINILYFVKTNDAIRV
                  ::  | : |::  || ::::|:  | |    : |: :|        |: ::  :: : :
253538a       LGEWQPIEYGKKKLKYLPYNHQHEYFFLIGPPLLIPMYFQYQI----IMTMIVHKNWVDL
              230       240       250       260       270            280

                   290       300       310       320       330       340
ma29gcg.pep   NPISTWHTVMFWGGKAFFVWYRLIVPLQYLPLGKVLLLFTVADMVSSYWLALTFQANHVV
                  :| : ::        ||: |     :|: |   || :||:::    :: |:|:: : | ||:|
253538a       ----AWAVSYYI---RFFITY---IPF-YGILG-ALLFLNFIRFLESHWFVWVTQMNHIV
                  290          300         310       320       330

                   350       360           370       380       390
ma29gcg.pep   EEVQWPLPDENGIIQKDWAAMQVETT----QDYAHDSHLWTSITGSLNYQAVHHLFPNVS
               |:     |:::   :|| : |: :|     |:: :|      | |   | ||:|  |||||::
253538a       MEI-----DQEAY--RDWFSSQLTATCNVEQSFFND---WFS--GHLNFQIEHHLFPTMP
                       340       350       360                370

               400       410       420       430       440
ma29gcg.pep   QHHYPDILAIIKNTCSEYKVPYLVKDTFWQAFASHLEHLRVLGLRPKEEX
              :|:    |   ::|: |::: : |   |
253538a       RHNLHKIAPLVKSLCAKHGIEYQEKPLLRALLDIIRSLKKSGKLWLDAYLHKX
            380       390       400       410       420       430
```

Figure 13

FastA Match of ma524 and contig 253538a

```
SCORES         Init1:   231 Initn:    499 Opt:    401
Smith-Waterman score: 620;     27.3% identity in 455 aa overlap

                    10        20        30        40        50       59
ma524gcg.pep MAAAPSVRTFTRAEVLNAEALNEGKKDAEAPFLMIIDNKVYDVREFVPDHPGGSVILTH-
                 |: | ||  ||         :::::    ::|| |||:: ||:  ||||| :::|
253538a          QGPTPRYFTWDEV--------AQRSGCEERWLVIDRKVYNISEFTRRHPGGSRVISHY
                         10                20        30        40        50

            60        70        80        90       100       110
ma524gcg.pep VGKDGTDVFDTFHPEAAW--ETLANFYVGDIDE---SDRDIKNDDFAAEVRKLRTLFQSL
             :|:|:|| | :|| : :   : : :: :|:::     | :  || ::: | |:||:  : :
253538a      AGQDATDPFVAFHINKGLVKKYMNSLLIGELSPEQPSFEPTKNKELTDEFRELRATVERM
                     60        70        80        90       100       110

                 120       130       140       150       160       170
ma524gcg.pep GYYDSSKAYYAFKVSFNLCIWGLSTVIVAKWGQTSTLANVLSAALLGLFWQQCGWLAHDF
             | : :::::: : :     | : | : : :  :| || |   :| |:||:     | ||| ||:
253538a      GLMKANHVFFLLYLLHILLLDGAAWLTLWVFG-TSFLPFLLCAVLLSAVQAQAGWLQHDY
                     120       130       140       150       160

                 180       190       200       210       220       230
ma524gcg.pep LHHQVFQDRFWGDLFGAFLGGVCQGFSSSWWKDKHNTHHAAPNVHGEDPDIDTHPLLTWS
              | :|::    |: |    |: |   :| |::||: :|  ||| ||:   :|||::    :|
253538a      GHLSVYRKPKWNHLVHKFVIGHLKGASANWWNHRHFQHHAKPNIFHKDPDVN---ML---
           170       180       190       200       210       220

                 240       250       260       270          280       290
ma524gcg.pep EHALEMFSDVPDEELTRMWSRFMVLNQTWFYFPILS---FARLSWCLQSILFVLPNGQAH
              |:: ::::      |   :   :::  |:     || :::   :   : :  | |: ::    :|
253538a      -HVF-VLGEWQPIEYGKKKLKYLPYNHQHEYFFLIGPPLLIPMYFQYQIIMTMI----VH
                  230       240       250       260       270

                    300       310       320          330       340      349
ma524gcg.pep KPSGARVPISLVEQLSLAMHWTWYLATMFLFIK--DPVNMLVYFLVSQAVCGNLLAIVFS
             |           : ::||   ::|:  :: :|         :: |:::   : : :: :: | :
253538a      K-----------NWVDLAWAVSYYIRFFITYIPFYGILGALLFLNFIRFLESHWFVWVTQ
                         280       290       300       310       320

           350       360       370       380       390       400      409
ma524gcg.pep LNHNGMPVISKEEAVDMDFFTKQIITGRDVHPGLFANWFTGGLNYQIEHHLFPSMPRHNF
             :||   | :     :||     |:|::|: :    :|: ::| :||:| ||:||||||||:|||||:
253538a      MNHIVMEI--DQEAYR-DWFSSQLTATCNVEQSFFNDWFSGHLNFQIEHHLFPTMPRHNL
              330       340       350       360       370       380

           410       420       430       440       450
ma524gcg.pep SKIQPAVETLCKKYNVRYHTTGMIEGTAEVFSRLNEVSKAASKMGKAQX
              || | |::|| |::::|:     ::::    :::   |:: :|
253538a      HKIAPLVKSLCAKHGIEYQEKPLLRALLDIIRSLKKSGKLWLDAYLHKX
                 390       400       410       420       430
```

Figure 14

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/07422

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/53 C12N15/82 C12N5/10 C12P7/64 C11B1/00
A61K31/20 A23L1/30 A23K1/00

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C12P C11B A61K A23L A23K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	COVELLO P. ET AL.: "Functional expression of the extraplastidial Arabidopsis thaliana oleate desaturase gene (FAD2) in Saccharomyces cerevisiae" PLANT PHYSIOLOGY, vol. 111, no. 1, May 1996, pages 223-226, XP002075211 see the whole document ---	1-83
A	WO 93 06712 A (RHONE POULENC AGROCHIMIE) 15 April 1993 cited in the application see the whole document ---	1-83
A	WO 94 18337 A (MONSANTO CO ;UNIV MICHIGAN (US); GIBSON SUSAN IRMA (US); KISHORE G) 18 August 1994 * see the whole document * ---	1-83

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 September 1998	21/09/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Kania, T

1

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 21022 A (RHONE POULENC AGROCHIMIE) 11 July 1996 cited in the application * see the whole document *	1-83
A	EP 0 561 569 A (LUBRIZOL CORP) 22 September 1993 cited in the application see the whole document	1-83
A	SPYCHALLA J. ET AL.: "Identification of an animal w3 fatty acid desaturase by heterologous expression in Arabidopsis" PNAS,U.S.A., vol. 94, no. 4, 18 February 1997, pages 1142-1147, XP002076628 * see esp. discussion *	1-83
A	HILLIER L. ET AL.: "The WashU-Merck EST Project, AC W49761" EMBL DATABASE, 30 May 1996, XP002076629 Heidelberg * corresponding to SEQ ID NO: 21 * see the whole document	86
A	NATHANS J.: "Adult human retina cDNA, AC W28140" EMBL DATABASE, 14 May 1996, XP002076630 Heidelberg * corresponding to SEQ ID NO: 22 * see the whole document	86
A	HILLIER L. ET AL.: "The WashU-Merck EST Project, AC W67716" EMBL DATABASE, 16 June 1996, XP002076631 Heidelberg * corresponding to SEQ ID NO: 24 * see the whole document	86
A	HILLIER L. ET AL.: "The WashU-Merck EST Project, AC H17219" EMBL DATABSE, 1 July 1995, XP002076632 Heidelberg * corresponding to SEQ ID NO: 25,27 * see the whole document	86

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	HILLIER L. ET AL.: "The WashU-Merck EST Project, AC H19385" EMBL DATABASE, 7 July 1995, XP002076633 Heidelberg * corresponding to SEQ ID NO: 26 * see the whole document	86
P,X	YOSHINOR. ET AL.: "AC C25549" EMBL DATABASE, 24 July 1997, XP002076634 Heidelberg * corresponding to SEQ ID NO: 13 * see the whole document	84,86-88
P,X	CADENA D. ET AL.: "AC AF002668" EMBL DATABASE, 4 July 1997, XP002076635 Heidelberg * corresponding to SEQ ID NO: 21 * see the whole document	86-88
T	MICHAELSON L. ET AL.: "Isolation of a delta5-fatty acid desaturase gene from Mortierella alpina" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 273, no. 30, 24 July 1998, pages 19055-19059, XP002076636 see the whole document	1-83

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 98/07422

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claims 57, 76, 77 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [X] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-83

   Nucleic acids, polypeptides, constructs comprising delta-5 desaturase according to SEQ ID NO: 1,2 derived from the fungus Mortierella alpina. Recombinant host cells comprising said nucleic acids or constructs.
   Methods for the production of arachidonic acid in a microbial cell comprising a cell containing a vector encoding an enzyme activity which converts dihomo-y-linolenic acid to arachidonic acid, preferentially a delta-5 desaturase, more preferentially from Mortierella.
   A recombinant yeast cell converting more than 5% of a 20:3 fatty acid to a 20:4 fatty acid.
   Methods for desaturating dihomo-y-linolenic acid using said microbial cells, fatty acids and oils obtained thereby.
   Methods for obtaining altered long chain polyunsaturated fatty acid biosynthesis using microbes comprising delta-5 desaturase derived from fungi or algae.
   Microbial oils derived from thereof and their use for therapeutical, nutritional, and cosmetical purposes, as well as products derived therefrom.

2. Claims: 84,86-88 partially

   An isolated sequence comprising the sequence of SEQ ID NO: 13, purified polypeptides produced thereof, esp. comprising SEQ ID NO: 14.

3. Claims: 84, 86-88 partially

   An isolated nucleotide comprising the sequence of SEQ ID NO: 15, purified polypeptides produced thereof, esp. comprising SEQ ID NO: 16.

4. Claims: 85 completely, 86-88 partially

   An isolated nucleotide sequence consisting of SEQ ID NO: 17 and SEQ ID NO: 19, purified polypeptides produced therefrom, esp. comprising SEQ ID NO: 18 and SEQ ID NO: 20.

./.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte ional Application No
PCT/US 98/07422

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9306712 A	15-04-1993	AU 667848 B	18-04-1996
		AU 2881292 A	03-05-1993
		BG 98695 A	31-05-1995
		BR 9206613 A	11-04-1995
		CA 2120629 A	15-04-1993
		CN 1072722 A	02-06-1993
		CN 1174236 A	25-02-1998
		CZ 9400817 A	13-09-1995
		EP 0666918 A	16-08-1995
		HU 69781 A	28-09-1995
		JP 7503605 T	20-04-1995
		MX 9205820 A	01-04-1993
		NZ 244685 A	27-06-1994
		US 5552306 A	03-09-1996
		US 5614393 A	25-03-1997
		US 5689050 A	18-11-1997
		US 5663068 A	02-09-1997
		US 5789220 A	04-08-1998
		ZA 9207777 A	21-04-1993
WO 9418337 A	18-08-1994	EP 0684998 A	06-12-1995
		JP 8506490 T	16-07-1996
WO 9621022 A	11-07-1996	US 5614393 A	25-03-1997
		AU 4673596 A	24-07-1996
		CA 2207906 A	11-07-1996
		CN 1177379 A	25-03-1998
		EP 0801680 A	22-10-1997
		US 5789220 A	04-08-1998
EP 0561569 A	22-09-1993	AU 3516793 A	16-09-1993
		CA 2092661 A	14-09-1993
		JP 6014667 A	25-01-1994
		US 5777201 A	07-07-1998

52
49

(19) Europäisches Pat ntamt
European Patent Offic
Offic européen d s brev ts

(11) Publication number: **0 249 676**
A2

(12) # EUROPEAN PATENT APPLICATION

(21) Application number: 87101137.5

(22) Date of filing: 28.01.87

(51) Int. Cl.4: C 12 N 15/00, C 12 P 21/02, C 12 N 1/20, C 12 N 5/00, A 01 H 1/00

(30) Priority: 28.01.86 DK 412/86

(43) Date of publication of application: 23.12.87 Bulletin 87/52

(84) Designated Contracting States: AT BE CH DE ES FR GB GR IT LI LU NL SE

(71) Applicant: AKTIESELSKABET DE DANSKE SUKKERFABRIKKER, Langebrogade 5, DK-1411 Copenhagen K (DK)

(72) Inventor: Marcker, Kjeld Adrian, 1, Toftevej, DK-8250 Egå (DK)
Inventor: Jensen, Jens Stougaard, 10, J.M. Morksgade 1. sal, DK-8000 Arhus C (DK)

(74) Representative: Vossius & Partner, Siebertstrasse 4 P.O. Box 86 07 67, D-8000 München 86 (DE)

(54) **Method for the expression of genes in plants.**

(57) A method for the expression of genes in plants, parts of plants, and plant cell cultures, in which a DNA fragment is used comprising an inducible plant promoter of root nodule-specific genes, DNA-fragments comprising an inducible plant promoter, to be used when carrying out the method, said DNA-fragments being identical with, derived from or comprising a 5' flanking region of root nodule-specific genes of any origin as well as plasmids and transformed Agrobacterium rhizogenes-strain which can be used when carrying out the method.

EP 0 249 676 A2

TITLE MODIFIED
see front page

0249676

<u>A method for the expression of genes in plants, parts of plants, and plant cell cultures, and DNA fragments, plasmids, and transformed microorganisms to be used when carrying out the method, as well as the use thereof for the expression of genes in plants, parts of plants, and plant cell cultures.</u>

The invention relates to a novel method for the expression of genes in plants, parts of plants, and plant cell cultures, as well as DNA fragments and plasmids comprising said DNA fragments to be used when carrying out the method. The invention furthermore relates to transformed plants, parts of plants and plant cells.

The invention relates to this method for the expression of genes of any origin under control of an inducible, root nodule specific promoter.

The invention relates especially to this method for the expression of root nodule-specific genes in transformed plants including both leguminous plants and other plants.

The invention relates furthermore to DNA fragments comprising an inducible plant promoter to be used when carrying out the method, as well as plasmids comprising said DNA fragments.

In the specification i.a. the following terms are used:

Root nodule-specific genes: Plant genes active only in the root nodules of leguminous plants, or

genes with an increased expression in root nodules. Root nodule-specific plant genes are expressed at predetermined stages of development and are activated in a coordinated manner during the symbiosis whereby a nitrogen fixation takes place and the fixed nitrogen is utilized in the metabolism of the plant.

Inducible plant promoter: Generally is meant a promoter-active 5' flanking region from plant genes inducible from a low activity to a high activity. In relation to the present invention "inducible plant promoter" means a promoter derived from, contained in or being identical with a 5' flanking region including a leader sequence of root nodule-specific genes and being capable of promoting and regulating the expression of a gene as characterised in relation to the present invention.

Leader sequence: Generally is meant a DNA sequence being transcribed into a mRNA, but not further translated into protein. The leader sequence comprises thus the DNA fragment from the start of the transcription to the ATG codon constituting the start of the translation. In relation to the present invention "leader sequence" means a short DNA fragment contained in the above inducible plant promoter and typically comprising 40-70 bp and which may comprise sequences being targets for a posttranscriptional regulation.

Promoter region: A DNA fragment containing a promoter which comprises target sequences for RNA polymerase as well as possible activation regions

comprising target sequences for transcriptional effector substances. In the present invention, target sequences for transcriptional effectors may also be situated 3′ to the promoter, i.e. in the coding sequences, the intervening sequences or on the 3′ flanking region of a root nodule-specific gene.

Furthermore a number of molecular-biological terms generally known to persons skilled in the art are used, including the terms stated below:

CAP (addition) site: The nucleotide of the 5′ end of the transcript where 7-methylGTP is added; In the Figures often given also as an asterisk *-marked nucleotide on a given nucleotide sequence.

DNA sequence or DNA segment: A linear array of nucleotides interconnected through phosphodiester bonds between the 3′ and 5′ carbon atoms of adjacent pentoses.

Expression: The process undergone by a structural gene to produce a polypeptide. It is a combination of transcription and translation as well as possible posttranslational modifications.

Flanking regions: DNA sequences surrounding coding regions. 5′ flanking regions contain a promoter. 3′ flanking regions may contain a transcriptional terminator etc.

Gene: A DNA sequence composed of three or four parts, viz. (1) the coding sequence for the gene

product, (2) the sequences in the promoter region which control whether or not the gene will be expressed, (3) those sequences in the 3' end conditioning the transcriptional termination and optionally polyadenylation, as well as (4) intervening sequences, if any.

Intervening sequences: DNA sequences within a gene which are not coding for any peptide fragment. The intervening sequences are transcribed into pre-mRNA and are eliminated by modification of pre-mRNA into mRNA. They are also called introns.

Chimeric gene: A gene composed of parts from various genes. E.g. the chimeric Lbc_3-5'-3'-CAT is composed of a chloroamphenicolacetyltransferase-coding sequence deriving from E. coli and 5' and 3' flanking regulatory regions of the Lbc_3 gene of soybean.

Cloning: The process of obtaining a population of organisms or DNA sequences deriving from one such organism or sequence by asexual reproduction, or more particular a process of isolating a particular organism or part thereof, and the propagation of this subfraction as a homogeneous population.

Coding sequences: DNA sequences determining the amino acid sequence of a polypeptide.

Cross-inoculation group: A group of leguminous plant species capable of producing functionally active root nodules with Rhizobium bacteria isolated from root nodules of other species of the group.

Leghemoglobin (Lb): An oxygen-binding protein exclusively synthesized in root nodules. The Lb proteins regulate the oxygen partial pressure in the root nodule tissue and transport oxygen to the bacteroides. In this manner the oxygen-sensitive nitrogenase enzyme is protected. The Lb genes are root nodule-specific genes.

Messenger-RNA (mRNA): RNA molecule produced by transcription of a gene and possibly modification of mRNA. The mRNA molecule mediates the genetic message determining the amino acid sequence of a polypeptide by part of the mRNA molecule being translated into said peptide.

Downstream: A position in a DNA sequence. It is defined relative to the transcriptional direction 5'- 3' of the gene relative to which the position is stated. The 3' flanking region is thus positioned downstream of the gene.

Nucleotide: A monomeric unit of DNA or RNA consisting of a sugar moiety (pentose), a phosphate, and a nitrogeneous heterocyclic base. The base is linked to the sugar moiety via a glycosidic bond (1' carbon of the pentose), and this combination of base and sugar is a nucleoside. The base characterises the nucleotide. The four DNA bases are adenine (A), guanine (G), cytosine (C), and thymine (T). The four RNA bases are A, G, C, and uracil (U).

Upstream: A position in a DNA sequence. It is defined relative to the transcriptional direction 5'- 3' f the gene relative to which the position

is stated. The 5' flanking region is thus positioned upstream of this gene.

Plant transformation: Processes leading to incorporation of genes in the genome of plant cells in such a manner that these genes are reliably inherited through mitosis and meiosis or in such a manner that these genes are only maintained for short periods.

Plasmid: An extra-chromosomal double-stranded DNA sequence comprising an intact replicon such that the plasmid is replicated in a host cell. When the plasmid is placed within a unicellular organism, the characteristics of that organism are changed or transformed as a result of the DNA of the plasmid. For instance a plasmid carrying the gene for tetracycline resistance (Tc^R) transforms a cell previously sensitive to tetracycline into one which is resistant to it. A cell transformed by a plasmid is called a transformant.

Polypeptide: A linear array of amino acids interconnected by means of peptide bonds between the α-amino and carboxy groups of adjacent amino acids.

Recombination: The creation of a new DNA molecule by combining DNA fragments of different origin.

Homologous recombination: A recombination between sequences showing a high degree of homology.

Replication: A process reproducing DNA molecules.

Replicon: A self-replicating genetic element possessing an origin for the initiation of DNA replication and genes specifying the functions necessary for a control and a replication thereof.

Restriction fragment: A DNA fragment resulting from double-stranded cleavage by an enzyme recognizing a specific target DNA sequence.

RNA polymerase: Enzyme effecting the transcription of DNA into RNA.

Root nodule: Specialized tissue resulting from infection of mainly roots of leguminous plants with Rhizobium bacteria. The tissue is produced by the host plant and comprises therefore plant cells whereas the Rhizobium bacteria upon infection are surrounded by a plant cell membrane and differentiate into bacteroides. Root nodules are produced on other species of plants upon infection of nitrogen-fixing bacteria not belonging to the Rhizobium genus. Root nodule-specific plant genes are also expressed in these nodules.

Southern-hybridization: Denatured DNA is transferred upon size separation in agarose gel to a nitrocellulose membrane. Transferred DNA is analysed for a predetermined DNA sequence or a predetermined gene by hybridization. This process allows a binding of single-stranded, radioactively marked DNA sequences (probes) to complementary single-stranded DNA sequences bound on the membrane. The position of DNA fragments on the membrane binding the probe can subsequently be detected n an X-ray film.

Symbiotic nitrogen fixation: The relationship whereby bacteroides of root nodules convert the nitrogen (dinitrogen) of the air into ammonium utilized by the plant while the plant provides the bacteroides with carbon compounds as a carbon source.

Symbiont: One part of a symbiotic relationship, and especially *Rhizobium* is called the microsymbiont.

Transformation: The process whereby a cell is incorporating a DNA molecule.

Translation: The process of producing a polypeptide from mRNA or:
the process whereby the genetic information present in a mRNA molecule directs the order of specific amino acids during the synthesis of a polypeptide.

Transcription: The method of synthesizing a complementary RNA sequence from a DNA sequence.

Vector: A plasmid, phage DNA or other DNA sequences capable of replication in a host cell and having one or a small number of endonuclease recognition sites at which such DNA sequences may be cleaved in a determinable manner without loss of an essential biological function.

Traditional plant breeding is based on repeated crossbreeding of plant lines individually carrying desired qualities. The identification of progeny lines carrying all the desired qualities is a particularly time-consuming process as the biochemical

and genetic basis of the qualities is usually unknown. New lines are therefore chosen according to their phenotype, usually after a screening of many lines in field experiments.

Through the ages a direct connection has existed between the state of nutrition, i.e. the health, of the population and the agricultural possibility of ensuring a sufficient supply of assimilable nitrogen in order to obtain satisfactory yields. Already in the seventeenth century it was discovered that plants of the family leguminosae including beyond peas also beans, lupins, soybean, bird's-foot trefoil, vetches, alfalfa, sainfoin, and trefoil had an ability of improving crops grown on the habitat of these plants. Today it is known that the latter is due to the fact that the members of the plants of the family leguminosae are able to produce nitrogen reserves themselves. On the roots they carry bacteria with which they live in symbiosis.

An infection of the roots of these leguminous plants with <u>Rhizobium</u> bacteria causes a formation of root nodules able to convert atmospheric nitrogen into bound nitrogen, which is a process called nitrogen fixation.

Atmospheric nitrogen is thereby converted into forms which can be utilized by the host plant as well as by the plants later on growing on the same habitat.

In the nineteenth century the above possibility was utilized f r the supply of nitrogen in order to achieve a novel increase of the crop yield.

The later further increases in the yield have, however, especially been obtained by means of natural fertilizers and nitrogen-containing synthetic fertilizers. The resulting pollution of the environment makes it desirable to provide alternative possibilities of ensuring the supply of nitrogen necessary for the best possible yields obtainable.

It would thus be valuable to make an improvement possible of the existing nitrogen fixation systems in leguminous plants as well as to allow an incorporation of nitrogen fixation systems in other plants.

The recombinant DNA technique and the plant transformation systems developed render it now possible to provide plants with new qualities in a well-controlled manner. These characteristics can derive from not only the same plant species, but also from all other prokaryotic or eukaryotic organisms. The DNA techniques allow further a quick and specific identification of progeny lines carrying the desired qualities. In this manner a specific plant line can be provided with one or more desired qualities in a quick and well-defined manner.

Correspondingly, plant cells can be provided with well defined qualities and subsequently be maintained as plant cell lines by means of known tissue culture methods. Such plant cells can be utilized for the production of chemical and biological products of particular interest such as dyes, flavours, aroma components, plant hormones, pharmaceutical

products, primary and secondary metabolites as well as polypeptides (enzymes).

A range of factors and functions necessary for biological production of a predetermined gene product are known. Both the initiation and regulation of transcription as well as the initiation and regulation of posttransscriptional processes can be characterised.

At the gene level it is known that these functions are mainly carried out by 5' flanking regions. A wide range of 5' flanking regions from prokaryotic and eukaryotic genes has been sequenced, and in view inter alia thereof a comprehensive knowledge has been provided of the regulation of gene expression and of the sub-regions and sequences being of importance for the regulation of expression of the gene. Great differences exist in the regulatory mechanism of prokaryotic and eukaryotic organisms, but many common features apply to the two groups.

The regulation of the expression of gene may take place on the transscriptional level and is then preferably exerted by regulating the initiation frequency of transscription. The latter is well-known and described inter alia by Benjamin Lewin, Gene Expression, John Wiley & Sons, vol. I, 1974, vol. II, Second Edition 1980, vol. III, 1977. As an alternative the regulation may be exerted at the posttransscriptional level, e.g. by the regulation of the frequency of the translation initiation, at the rate of the translation, and of the termination of the translation.

The present invention is based on the surprising finding that 5' flanking regions of root nodule-specific genes, exemplified by the 5' flanking region of the soybean leghemoglobin Lbc_3 gene, can be used for inducible expression of a foreign gene in an alien leguminous plant. The induction and regulation of the promoter is preferably carried out in the form of a regulation and induction at the transscriptional level and differs thereby from the inducability stated in Patent Application No. 86114704.9, the latter inducability preferably being carried out at the translation level.

The transscription of both the Lbc_3 gene of the soybean and of a chimeric Lbc_3 gene transferred to bird's-foot trefoil starts at a low level immediately upon the appearance of the root nodules on the plant roots. Subsequently, a high increase of the transscription takes place immediately before the root nodules turn red. The transcription of a range of other root nodule-specific genes is initiated exactly at this time. The simultaneous induction of the transscription of the Lb genes and other root nodule-specific genes means that a common DNA sequence(s) must be present for the various genes controlling this pattern of expression. Thus the leghemoglobin-c_3 gene is a representative of one class of genes and the promoter and the leader sequence, target areas for activation as well as the control elements of the organ specificity of the Lbc_3 gene are representatives of the control elements of a complete gene class.

The promoter of the 5' flanking regions of the Lb genes functions in soybeans and is responsible for the transcription of the Lb genes in root nodules. It is furthermore known, that the efficiency of both the transcription initiation and the subsequent translation initation on the leader sequence of the Lb genes is high as the Lb proteins constitute approximately 20% of the total protein content in root nodules.

The sequence of 5' flanking regions of the four soybean leghemoglobin genes Lba, Lbc_1, Lbc_2, and Lbc_3 appears from the enclosed sequence scheme, scheme 1, wherein the sequences are stated in such a manner that the homology between the four 5' flanking regions appears clearly.

In the sequence scheme "-" indicates that no base is present in the position in question. The names of the genes and the base position counted upstream from the ATG start codon are indicated to the right of the sequence scheme. Furthermore the important sequences have been underlined.

As it appears from the sequence scheme a distinct degree of homology exists between the four 5' flanking regions, and in the position 23-24 bp upstream from the CAP addition site they all contain a TATATAAA sequence corresponding to the "TATA" box which in eukaryotic cells usually are located a corresponding number of bp upstream from the CAP addition site. Furthermore a CCAAG sequence is present 64-72 bp upstream from the CAP addition site, said sequence corresponding to the "CCAAT"

box usually located 70-90 bp upstream from the CAP addition site. From the CAP addition site to the translation start codon, ATG, leader sequences of 52-59 bp are present and show a distinct degree of homology of approx. 75-80%.

In accordance with the present invention it has furthermore been proved, exemplified by the Lbc_3 gene, that the 5' flanking regions of the soybean leghemoglobin genes are functionally active in other plant species. The latter has been proved by fusioning the E. coli chloroamphenicol acetyl transferase (CAT) gene with the 5' and 3' flanking regions of the soybean Lbc_3 gene in such a manner that the expression of the CAT gene is controlled by the Lb promoter. This fusion fragment was cloned into the integration vectors pAR1 and pAR22, whereby the plasmids pAR29 and pAR30 were produced. Through homologous recombination the latter plasmids were integrated into the Agrobacterium rhizogenes T DNA region. The transformation of Lotus corniculatus (bird's-foot trefoil) plants, i.e. transfer of the T DNA region, was obtained by wound infection on the hypokotyl. Roots developed from the transformed plant cells were cultivated in vitro and freed from A. rhizogenes bacteria by means of antibiotics. Completely regenerated plants were produced by these root cultures in a conventional manner through somatic embryogenesis or organogenesis.

Regenerated plants were subsequently inoculated with Rhizobium loti bacteria and root nodules for analysis were harvested. Transcription and translation of the chimeric Lbc_3 CAT gene could subse-

quently be detected in root nodules on transformed plants as the activity of the produced chloroamphenicol acetyl transferase enzyme.

The conclusion can subsequently be made that the promoter-containing 5' flanking regions of root nodule-specific genes exemplified by the soybean Lbc_3 promoter are functionally active in foreign plants. The latter is a surprising observation as root nodules are only developed as a consequence of a very specific interaction between the leguminous plant and its corresponding <u>Rhizobium</u> microsymbiont.

Soybeans produce nodules only upon infection by the species <u>Rhizobium japonicum</u> and <u>Lotus corniculatus</u> only upon infection by the species <u>Rhizobium loti</u>. Soybean and <u>Lotus corniculatus</u> belong therefore to two different cross-inoculation groups, each group producing root nodules by means of two different <u>Rhizobium</u> species. The expression of a chimeric soybean gene in <u>Lotus corniculatus</u> proves therefore an unexpected universal regulatory system applying to the expression of root nodule-specific genes. The regulatory DNA sequences involved can be placed on the 5' and 3' flanking regions of the genes, here exemplified by the 2.0 Kb 5' and 0.9 Kb 3' flanking regions of the Lbc_3 gene. This surprising observation allows the use of root nodule-specific promoters and regulatory sequences in any other plant species and any other plant cell line.

In other experiments the 5' flanking region of the nodule-specific N23 gene was fused to the CAT gene

and the Lbc_3 3' flanking region in such a manner that the expression of the CAT gene is controlled by the N23 promoter. This fusion fragment was cloned into the integration vector pAR22 producing the plasmid N23-CAT which was subsequently recombined into A.rhizogenes and transferred to Lotus corniculatus and Trifolium repens (white clover) by the previously described method. The root nodule-specific expression of the transferred N23-CAT gene obtained in L. corniculatus infected with Rhizobium loti and in T.repens infected with Rhizobium trifolii further demonstrated that expression of root nodule-specific genes is independent of the plant species and Rhizobium species. A universal regulatory system therefore regulates the expression of root nodule-specific genes in the different symbiotic systems formed between legumes and the Rhizobium species of the various cross-inoculation groups.

It is known from European Patent Application EP 122,791.A1 that plant genes from one species, by Agrobacterium mediated transformation, can be transferred into a different plant species. It is also known from EP 122,791.A1 that a transferred gene encoding the seed storage protein "Phaseolin" can be expressed into tobacco and alfalfa. From the literature it is also known that this expression is seed specific (Sengupta-Gopalan et al. 1985, Proc. Natl. Acad. Sci. 82, 33203324).

The present invention therefore relates to a novel method for the expression of transferred genes in a root nodule-specific manner, using DNA regulatory

sequences from the 5' promoter region, the coding region, or the 3' flanking region of root nodule-specific genes, here exemplified by the leghemoglobin Lbc_3 gene and the N23 gene. This method is distinct from both the method of Agrobacterium mediated transformation and expression of the seed storage protein phaseolin gene characterised in EP 122,791.A1. Expression of the transferred phaseolin gene in EP 122,791.A1 only demonstrates that the phaseolin gene family with its particular regulatory requirements can be expressed in tobacco and alfalfa. It does not demonstrate nor predict that any other genes with their particular regulatory requirements can be expressed in any other plants or plant tissue.

An object of the present invention is to provide a possibility of expressing desired genes in plants, parts of plants, and plant cell cultures.

A further object of the invention is to render it possible to express genes of any origin by the control of an inducible root nodule-specific promoter.

A particular object of the invention is to provide a possibility of expressing desired genes in leguminous plants.

A still further particular object of the invention is to provide a possibility of expressing root nodule-specific genes in non-leguminous plants.

Further objects of the invention are to improve the

existing nitrogen-fixing systems in leguminous plants as well as to incorporate nitrogen-fixing systems in other plants.

A further object of the invention is to provide a possibility of in certain cases allowing the use of specific sequences of the 3' flanking region, of the coding sequence, and of intervening sequences to influence the regulation of the root nodule-specific promoter.

Furthermore it is an object of the invention to provide plasmids comprising the above mentioned inducible plant promoter.

Further objects of the invention appear immediately from the following description.

The method according to the invention for the expression of genes in plants, parts of plants, and plant cell cultures is carried out by introducing into a cell thereof a recombinant DNA segment containing both the gene to be expressed and a 5' flanking region comprising a promoter sequence, and optionally a 3' flanking region, and culturing of the transformed cells in a growth medium, said method being characterised by using as the recombinant DNA segment a DNA fragment comprising an inducible plant promoter (as defined) from root nodule-specific genes. If desired the transformed cells are regenerated to plants.

The method according to the invention allows in a well defined manner an expression f foreign genes

in plants, parts of plants, and plant cell cultures, in this connection especially genes providing the plants with desired properties such as for instance a resistance to plant diseases and increased content of valuable polypeptides.

A further use is the preparation of valuable products such as for instance dyes, flavourings, plant hormones, pharmaceutical products, primary and secondary metabolites, and polypeptides by means of the method according to the invention in plant cell cultures and plants.

By using the method according to the invention for the expression of root nodule-specific genes it is possible to express root nodule-specific genes necessary for the formation of an active nitrogen-fixing system both in leguminous plants and other plants. The correct developmental control, cf. Example 8, allows the establishment of a symbiotic nitrogen-fixing system in non-leguminous plants. In this manner it is surprisingly possible to improve the existing nitrogen-fixing systems in leguminous plants as well as to incorporate nitrogen-fixing systems in other plants.

The use of the method according to the invention for the expression of foreign genes in root nodules renders it possible to provide leguminous plants with improved properties such as resistance to herbicides and resistance to diseases and pest.

According to a particular embodiment of the method according to the invention a DNA fragm nt is used

which comprises an inducible plant promoter and which is identical with, derived from, or comprises 5′ flanking regions of leghemoglobins genes. In this manner the expression of any gene is obtained.

Examples of such DNA fragments are DNA fragments of the four 5′ flanking regions of the soybean leghemoglobin genes, viz.

Lba with the sequence:

GAGATACATT ATAATAATCT CTCTAGTGTC TATTTATTAT TTTATCTGGT
GATATATACC TTCTCGTATA CTGTTATTTT TTCAATCTTG TAGATTTACT
TCTTTTATTT TTATAAAAAA GACTTTATTT TTTTAAAAAA AATAAAGTGA
ATTTTGAAAA CATGCTCTTT GACAATTTTC TGTTTCCTTT TTCATCATTG
GGTTAAATCT CATAGTGCCT CTATTCAATA ATTTGGGCTC AATTTAATTA
GTAGAGTCTA CATAAAATTT ACCTTAATAG TAGAGAATAG AGAGTCTTGG
AAAGTTGGTT TTTCTCGAGG AAGAAAGGAA ATGTTAAAAA CTGTGATATT
TTTTTTTTGG ATTAATAGTT ATGTTTATAT GAAAACTGAA AATAAATAAA
CTAACCATAT TAAATTTAGA ACAACACTTC AATTATTTTT TTAATTTGAT
TAATTAAAAA ATTATTTGAT TAAATTTTTT AAAAGATCGT TGTTTCTTCT
TCATCATGCT GATTGACACC CTCCACAAGC CAAGAGAAAC ACATAAGCTT
TGGTTTTCTC ACTCTCCAAG CCCTCTATAT AAACAAATAT TGGAGTGAAG
TTGTTGCATA ACTTGCATCG AACAATTAAT AGAAATAACA GAAAATTAAA
AAAGAAATAT G,

Lbc$_1$ with the sequence:

TTCTCTTAAT ACAATGGAGT TTTTGTTGAA CATACATACA TTTAAAAAAA
AATCTCTAGT GTCTATTTAC CCGGTGAGAA GCCTTCTCGT GTTTTACACA
CTTTAATATT ATTATATCCT CAACCCCACA AAAAAGAATA CTGTTATATC
TTTCCAAACC TGTAGATTTA TTTATTTATT TATTTATTTT TACAAAGGAG
ACTTCAGAAA AGTAATTACA TAAAGATAGT GAACATCATT TTATTTATTA
TAATAAACTT TAAAATCAAA CTTTTTTATA TTTTTTGTTA CCCTTTTCAT
TATTGGGTGA AATCTCATAG TGAAGCCATT AAATAATTTG GGCTCAAGTT
TTATTAGTAA AGTCTGCATG AAATTTAACT TAACAATAGA GAGAGTTTTC
GAAAGGGAGC GAATGTTAAA AAGTGTGATA TTATATTTTA TTTCGATTAA
TAATTATGTT TACATGAAAA CATACAAAAA AATACTTTTA AATTCAGAAT
AATACTTAAA ATATTTATTT GCTTAATTGA TTAACTGAAA ATTATTTGAT
TAGGATTTTG AAAAGATCAT TGGCTCTTCG TCATGCCGAT TGACACCCTC
CACAAGCCAA GAGAAACTTA AGTTGTAAAC TTTCTCACTC CAAGCCTTCT
ATATAAACAT GTATTGGATG TGAAGTTATT GCATAACTTG CATTGAACAA
TAGAAAATAA CAAAAAAAAG TAAAAAAGTA GAAAAGAAAT ATG,

Lbc_2 with the sequence:

```
TCGAGTTTTT ACTGAACATA CATTTATTAA AAAAAACTCT CTAGTGTCCA
TTTATTCGGC GAGAAGCCTT CTCGTGCTTT ACACACTTTA ATATTATTAT
ATCCCCACCC CCACCAAAAA AAAAAAAACT GTTATATCTT TCCAGTACAT
TTATTTCTTA TTTTTACAAA GGAAACTTCA CGAAAGTAAT TACAAAAAAG
ATAGTGAACA TCATTTTTTT AGTTAAGATG AATTTTAAAA TCACACTTTT
TTATATTTTT TTGTTACCCT TTTCATTATT GGGTGAAATC TCATAGTGAA
ACTATTAAAT AGTTTGGGCT CAAGTTTTAT TAGTAAAGTC TGCATGAAAT
TTAACTTAAT AATAGAGAGA GTTTTGGAAA GGTAACGAAT GTTAGAAAGT
GTGATATTAT TATAGTTTTA TTTAGATTAA TAATTATGTT TACATGAAAA
TTGACAATTT ATTTTTAAAA TTCAGAGTAA TACTTAAATT ACTTATTTAC
TTTAAGATTT TGAAAAGATC ATTTGGCTCT TCATCATGCC GATTGACACC
CTCCACAAGC CAAGAGAAAC TTAAGTTGTA ATTTTTCTAA CTCCAAGCCT
TCTATATAAA CACGTATTGG ATGTGAAGTT GTTGCATAAC TTGCATTGAA
CAATAGAAAT AACAACAAAG AAAATAAGTG AAAAAAGAAA TATG,
```

and Lbc_3 with the sequence:

```
TATGAAGATT AAAAAATACA CTCATATATA TGCCATAAGA ACCAACAAAA
GTACTATTTA AGAAAAGAAA AAAAAAACCT GCTACATAAT TTCCAATCTT
GTAGATTTAT TTCTTTTATT TTTATAAAGG AGAGTTAAAA AAATTACAAA
ATAAAAATAG TGAACATCGT CTAAGCATTT TTATATAAGA TGAATTTTAA
AAATATAATT TTTTTGTCTA AATCGTATGT ATCTTGTCTT AGAGCCATTT
TTGTTTAAAT TGGATAAGAT CACACTATAA AGTTCTTCCT CCGAGTTTGA
TATAAAAAAA ATTGTTTCCC TTTTGATTAT TGGATAAAAT CTCGTAGTGA
CATTATATTA AAAAAATTAG GGCTCAATTT TTATTAGTAT AGTTTGCATA
AATTTTAACT TAAAAATAGA GAAAATCTGG AAAAGGGACT GTTAAAAAGT
GTGATATTAG AAATTTGTCG GATATATTAA TATTTTATTT TATATGGAAA
CTAAAAAAAT ATATATTAAA ATTTTAAATT CAGAATAATA CTTAAATTAT
TTATTTACTG AAAATGAGTT GATTTAAGTT TTTGAAAAGA TGATTGTCTC
TTCACCATAC CAATTGATCA CCCTCCTCCA ACAAGCCAAG AGAGACATAA
GTTTTATTAG TTATTCTGAT CACTCTTCAA GCCTTCTATA TAAATAAGTA
TTGGATGTGA AGTTGTTGCA TAACTTGCAT TGAACAATTA ATAGAAATAA
CAGAAAAGTA GAAAAGAAAT ATG.
```

A further embodiment of the method according to the invention uses a DNA fragment identical with, derived from or comprising 5' flanking regions of the Lbc_3-5'-3'-CAT gene with the sequence:

```
TATGAAGATT AAAAAATACA CTCATATATA TGCCATAAGA ACCAACAAAA
GTACTATTTA AGAAAAGAAA AAAAAAACCT GCTACATAAT TTCCAATCTT
GTAGATTTAT TTCTTTTATT TTTATAAAGG AGAGTTAAAA AAATTACAAA
ATAAAAATAG TGAACATCGT CTAAGCATTT TTATATAAGA TGAATTTTAA
AAATATAATT TTTTTGTCTA AATCGTATGT ATCTTGTCTT AGAGCCATTT
TTGTTTAAAT TGGATAAGAT CACACTATAA AGTTCTTCCT CCGAGTTTGA
```

```
TATAAAAAAA ATTGTTTCCC TTTTGATTAT TGGATAAAAT CTCGTAGTGA
CATTATATTA AAAAAATTAG GGCTCAATTT TTATTAGTAT AGTTTGCATA
AATTTTAACT TAAAAATAGA GAAAATCTGG AAAAGGGACT GTTAAAAAGT
GTGATATTAG AAATTTGTCG GATATATTAA TATTTTATTT TATATGGAAA
CTAAAAAAAT ATATATTAAA ATTTTAAATT CAGAATAATA CTTAAATTAT
TTATTTACTG AAAATGAGTT GATTTAAGTT TTTGAAAAGA TGATTGTCTC
TTCACCATAC CAATTGATCA CCCTCCTCCA ACAAGCCAAG AGAGACATAA
GTTTTATTAG TTATTCTGAT CACTCTTCAA GCCTTCTATA TAAATAAGTA
TTGGATGTGA AGTTGTTGCA TAACTTGCAT TGAACAATTA ATAGAAATAA
CAGAAAAGTA GAATTCTAAA ATG
```

A still further preferred embodiment of the method according to the invention uses a DNA fragment identical with, derived from or comprising 5' flanking regions of the N23 gene with the sequence

```
         10        20        30        40        50        60        70
GAATTCGAGCTCGCCCGGGGATCGATCCTCTAGAGTCGACCTGCAGCCCAAGCTTGGATCAATCAATTAA
EcoRI                              SalI

         80        90       100       110       120       130       140
TTCTATTGAGACACGATTTGAACAATTTTTACATTATGAGACTATTTTTGGTTTTTTATTTGATCCAAAA

        150       160       170       180       190       200       210
AAATTTAAAGCTTTAGATGATGATGAATTGAANNAATATTGTATTAATNNTGAAAAGTTNNNNNGGTTTA

        220       230       240       250       260       270       280
ATGAATGCTATGATATTGATGGTCTTGATNTATTNNCAGAATTGAAAGTATTAAGAGAAGTGTTAAGAAA

        290       300       310       320       330       340       350
AGAAGTTAGCACACCAATAGAAGTATTGAGTTATATTAAAACTTTAGATTCTTTTCAAATGTTTACATTG

        360       370       380       390       400       410       420
CATATAGAATTTTATTGACAATCCTTATAACAGTTGCTACTGTTGAAAGACGTTCTTCAAAATTAAAATT

        430       440       450       460       470       480       490
ACTTAAATCATATCTAAAATCAACAATGTTACAAGATAGATTGAATGAGTTAGTTATTTTATCTATTGAA

        500       510       520       530       540       550       560
AGTAAAGTGTTAGAATTGTTTGATTATAAAACTCTGATAAATGATTTTGCAGTTAAAAAAACTAGAAGAT

        570       580       590       600       610       620       630
TAATATAAAAATTGATATTTTATATAATATATTAAGTCTCTTTAAAATTCTTGTAAAAAAAGACATTTTT

        640       650       660       670       680       690       700
AAATAATAAAATAAAGCAACTCTTAATTTTAATGAAACATCCCTTTGTTAAACCGAATCTTCCATAATGT

        710       720       730       740       750       760       770
AAAAATTAATGCTTGATGGAAGTTTTTAATTTGTTCTACTCAATACTCAAAGGGTTGTAAATATTTTTTT

        780       790       800       810       820       830       840
TATCATTTATATGTTGTAAATATGAATGCACTAGTAATTAGTTTAATGATAAAATATATTCTACAGATAT
```

```
        850       860       870       880       890       900       910
ATTTCTGTCTCTTGGCAACTCGTGAGAATTGAATATATTATAAAGATGAAAGGTCGTTACAATTTTTTTT

        920       930       940       950       960       970       980
AGAATAAATATTTATATACAATTCCTAGATTTTGTTATAAAATTCACATATTGTATGAGTATAAATACAT

        990      1000      1010      1020      1030      1040      1050
GAGCACACACCAAACTAGTCTCAAATTAAGTAAGGTGCTAATTATTAGCGGCTAGCTAAGTAACCAAGTA
                                                        DdeI

ATTAATG
```

In a particularly preferred embodiment of the method according to the invention a 3' flanking region of root nodule-specific genes is furthermore used, in particular sequences of the 3' flanking region capable of influencing the activity or regulation of a promotor of the root nodule-specific genes or the transcription termination, or capable of influencing the yield of the desired gene product in another manner.

Examples of such 3' flanking regions are the four 3' flanking regions of the soybean leghemoglobin genes, viz.

Lba with the sequence:

```
                                        1590                                     1620
                    TAA TTA GTA TCT ATT GCA GTA AAG TGT AAT AAA TAA ATC TTG

                                        1650                                     1680
TTT CAC TAT AAA ACT TGT TAC TAT TAG ACA AGG GCC TGA TAC AAA ATG TTG GTT AAA ATA

                                        1710                                     1740
ATG GAA TTA TAT AGT ATT GGA TAA AAA TCT TAA GGT TAA TAT TCT ATA TTT GCG TAG GTT

                                        1770                                     1800
TAT GCT TGT GAA TCA TTA TCG GTA TTT TTT TTC CTT TCT GAT AAT TAA TCG GTA AAT TA

                                        1830                                     1860
ACA AAT AAG TTC AAA ATG ATT TAT ATG TTT CAA AAT TAT TTT AAC AGC AGG TAA AAT GTT

ATT TGG TAC GAA AGC TAA TTC GTC GA
```

Lbc$_1$ with the sequence:

```
                                                                      1320
                                        TAA/TT AGG ATC TAC TGC ATT GCC GTA

                                     1350                                 1380
AAG TGT AAT AAA TAA ATC TTG TTT CAA CTA AAA CTT GTT ATT AAA CAA GTT CCC TAT ATA

                                     1410                                 1440
AAT GTT GTT TAA AAT AAG TAA ATT TCA TTG TAT TGG ATA AAC ACT TTT AAG TTA TAT ATT

                                     1470                                 1500
TCC ATA TAT TTA CGT TTG TGA ATC ATA ATC GAT ACT TTA TAA AAA TAA ATT CCA AAT AAT

TTA TAC GTT TTA AAA ATT ATT TT
```

Lbc$_2$ with the sequence:

```
                                            TAG/GAT CTA CTA TTG CCG TCA AGT
                                                                       1140

GTA ATA AAT AAA TTT TGT TTC ACT AAA ACT TGT TAT TAA ACA AGT CCC CGA TAT ATA AAT
                                     1170                                 1200
GTT GGT TAA AAT AAG TAA ATT ATA CGG TAT TGA TAA ACA ATC TTA AGT TTT ATA TAT AGT
                                     1230                                 1260

TCC ATA TAC TAA AGT TTG TGA ATC ATA ATC GA
                                     1290
```

and Lbc$_3$ with the sequence:

```
                            TAG/GAT CTA CAA TTG CCT TAA AGT GTA ATA AAT AAA
                                 990                                  1020

TAT TAT TTC ACT AAA ACT TGT TAT TAA ACC AAG TTC TCG ATA TAA ATG TTG GTT AAA CTA
                                     1050                                 1080

AGT AAA TTA TAT GGT ATT GGA TAA ACA ATC TTA AGC TT
                                     1110
```

This sequence is positioned on the 0.9 Kb 3′ flanking region used according to the invention. A particular embodiment of the invention is therefore

the use of sequences of this region exerting or mediating the regulation characterised by the invention of root nodule-specific promoter regions.

In a preferred embodiment of the method according to the invention a region is used of the coding sequence or intervening sequence of root nodule-specific genes, in particular sequences of the coding sequence or the intervening sequence capable of influencing the regulation of a promotor of the root nodule-specific genes or capable of influencing the yield of the desired gene product in another manner.

Examples of such coding sequences and intervening sequences are the four leghemoglobin genes of soybean, viz.

Lba with the sequence:

```
                                                                      120
                                                                      VAL
                                                                  ATG/GTT

                                      150                             180
ALA PHE THR GLU LYS GLN ASP ALA LEU VAL SER SER SER PHE GLU ALA PHE LYS ALA ASN
GCT TTC ACT GAG AAG CAA GAT GCT TTG GTG AGT AGC TCA TTC GAA GCA TTC AAG GCA AAC

                                      210                             240
ILE PRO GLN TYR SER VAL VAL PHE TYR THR SER
ATT CCT CAA TAC AGC GTT GTG TTC TAC ACT TC/G TAA GTT TTC TCT CTA AGC ATG TGT CTT

                                      270                             300
CCA TTC TAT GTT TTT CTT TTG GAA ATT TGT TGT GTT TGA AAA AAG ATA TAT TGT TAA TGT

                                      330                              360
                                          ILE LEU GLU LYS ALA PRO ALA ALA LYS ASP
GAG TGG TTT TGG TTT GAT TAA AAA TGA ATAG/G ATA CTG GAG AAA GCA CCT GCA GCA AAG GAC

                                  390                                 420
LEU PHE SER PHE LEU ALA ASN GLY VAL ASP PRO THR ASN PRO LYS LEU THR GLY HIS ALA
TTG TTC TCA TTT CTA GCA AAT GGA GTA GAC CCC ACT AAT CCT AAC CTC ACG GGC CAT GCT

                                  450                               480
GLU LYS LEU PHE ALA LEU
GAA AAG CTT TTT GCA TTG/GTAA GTA TCA CCC AAC TAA AAT TAT AAC TAT TTT ATG TGA

                                      510                             540
TTA ATT TTA AGA TTA AGC ATC ATG TAT TTT AAC ACT CTT AAA ACA TCA ATG AAC ATT AAT

                                      570                             600
TGT TTG AAT TGT ATT TTA TAT TTT TGC CAT ATC TTG AAC TAG GAA TAG TAT ATA AAT TTC

                                      630                             660
TAT TAG TAT TTG TTG ATA ATT ATT TTT CTT TCA TAA CTA TCT TGT CAC ATA TTA TAT ATT
```

690 720
VAL ARG ASP SER ALA GLY GLN LEU LYS ALA SER GLY THR VAL VAL ALA
TTT TGA ATT GTAG/GTG CGT GAC TCA GCT GGT CAA CTT AAA GCA AGT GGA ACA GTG GTG GCT

750 780
ASP ALA ALA LEU GLY SER VAL HIS ALA GLN LYS ALA VAL THR ASP PRO GLN PHE VAL
GAT GCC GCA CTT GGT TCT GTT CAT GCC CAA AAA GCA GTC ACT GAT CCT CAG TTC GTG/GT

810 840
ATG ATA AAT AAT GAA ATG TTA TAA TAA ATT ATG CAT ACT TCA ATT TTT CAT GGA GCA GTA

870 900
TAA TGA TCA ACA CAC ACT TCT TTT GTT TCA TGC ATT TGA TAA CTA CAA TCT TAA AAT GTT

930 960
GCA ATC TTA AAA ATA GTA TTA AAA ATA TAA CAT TTA ATT AGC TCA TCA ATA TTT TTC TGT

990 1020
TGC AAT TTT TTA TGA AAA AAT TAT AAT TAT GAA TTC TTT GAG CAA TGT TTA ATT AAA AAA

1050 1080
TTG ATT TAA TAA TGA AAT AAC TAA GCT ACC TCT GTC TCG TTT TTC ATT TAA ACT ATG ACA

1110 1140
TAA ACA ATG AAT AAA GTA AAC TAA ACC ATG ACA TGT TTA TTT TTG AAT GAG GTT ATT AAT

1170 1200
AAT TTT TTT TCA CTA TCT ATT GCA ATG TTC ATT GAT TAT CAA TTA TCT TGG TTG CAT TGA

1230 1260
TTC TCT CGA TTT TTT TCT TGA GGT TAA GCT TCA GTT CAA TAT ATA TTC ATT TTT TGA TAA

1290 1320
AAA AAA ATA GTA CAA TAT ATT TTC ATT TAG CTG ATC ATA TTT ATT TAA GTT CAA CTT AAA

1350 1380
ATT TTA TAG ATG TTA ATT GAT ATA ATT TGT TGA GAT GAT GAG AAG ACC AAT ACC ATT ACG

1410 1440
TAC TCT TTT GAA AGT GTT ATA TGG ATT TTA ATT ATA AGG AAA AAT GTA AGA GCT AAA CCA

1470 1500
VAL VAL LYS GLU ALA LEU LEU LYS THR ILE LYS ALA ALA VAL
TTG CTG ATG ATT TTG AAG/GTG GTT AAA GAA GCA CTG CTG AAA ACA ATA AAG GCA GCA GTT

1530 1560
GLY ASP LYS TRP SER ASP GLU LEU SER ARG ALA TRP GLU VAL ALA TYR ASP GLU LEU ALA
GGG GAC AAA TGG AGT GAC GAG TTG AGC CGT GCT TGG GAA GTA GCC TAC GAT GAA TTG GCA

ALA ALA ILE LYS LYS ALA
GCA GCT ATT AAG AAG GCA TAA

The amino acid sequence of the Lba protein is indicated above the coding sequence,

Lbc_1 with the sequence:

```
                                                                  180
                                                                  GLY
                                                              ATG/GCT

                                   210                                      240
ALA PHE THR GLU LYS GLN GLU ALA LEU VAL SER SER SER PHE GLU ALA PHE LYS ALA ASN
GCT TTC ACT GAG AAG CAA GAG GCT TTG GTG AGT AGC TCA TTC GAA GCA TTC AAG GCA AAC

                                   270                                      300
ILE PRO GLN TYR SER VAL VAL PHE TYR ASN SER
ATT CCT CAA TAC AGC GTT GTG TTC TAC AAT TC/GTAA GTT TTC TCT ATA AGC ATG TGT CTT

                                   330                                      360
TCA TTC TAT GTT TTT CTT CTG GAA ATT TTT TGT GTT TGA AAA AAG ATA TAT ATA TAT ATA

                                   390                                      420
TAT ATA TAT ATA TAT ATA TAT ATA TAT ATA TAT ATA TAT TTT GTT AAT GTG AGT GGT TTT

                                   450                                      480
                             ILE LEU GLU LYS ALA PRO ALA ALA LYS ASP LEU PHE SER
GGT TTG ATT AAA AAT AAA TAG/GATT CTG GAG AAA GCA CCT GCA GCA AAG GAC TTG TTC TCA

                                   510                                      540
PHE LEU ALA ASN GLY VAL ASP PRO THR ASN PRO LYS LEU THR GLY HIS ALA GLU LYS LEU
TTT CTA GCA AAT GGA GTA GAC CCC ACT AAT CCT AAG CTC ACG GGC CAT GCT GAA AAG CTT

                               570                                         600
PHE ALA LEU
TTT GCA TTG/GT AAG TAT CAG CCA ACT AAA ATT ATA ACT ATT TTA TGT GAT TAA TTT TAA

                                   630                                      660
GAT TAA ACA TCA TGT ATT TTA ACA CTC TTA AAA TAT CAA TGA ACA TTA ATT TTT TGA ATT

                                   690                                      720
GTA TTT TAT ATT TTT ACC ATA TCT TGA ACT AGG AAT AAT ATA TAA ATT TCT ATT AGT ATT

                                   750                                      780
TGT TGG TAA TTA CAT ATA TAT ATA TAT ATA TAA TCC TTG TGA TAA TTA TTT TTC GAA TTT

                                   810                                      840
     VAL ARG ASP SER ALA GLY GLN LEU LYS THR ASN GLY THR VAL VAL ALA ASP ALA ALA
GTAG/GTG CGT GAC TCA GCT GGT CAA CTT AAA ACA AAT GGA ACA GTG GTG GCT GAT GCT GCA

                                   870                                      900
LEU VAL SER ILE HIS ALA GLN LYS ALA VAL THR ASP PRO GLN PHE VAL
CTT GTT TCT ATC CAT GCC CAA AAA GCA GTC ACT GAT CCT CAG TTC GTG/GT ATG ATA AAT

                                   930                                      960
AAT ACT AGT AAA ATG TTA CAA TAA ATG CAA ACT TAA GTT TTA CGT ACA TAG TGA TCA TGA

                                   990                                     1020
CTT CAT GCA TGG CTA TTA TTT TTT CAT ATT TAT TGA AGT CAA CTT AAA ATT TTG TAA ATA

                                  1050                                     1080
CAG ATC GAT GCT AGT AAT TTG TTG AGA TCA TGA GAA AAC GTA CCA CTA CTC CAA TAG CAT

                                  1110                                     1140
TAC TCA TTT TGA AAA TTG TAT AAC TGT GAT CTA ATT ATA AGG AAA AAG TGT ATA TAA GAG

                                  1170                                     1200
                                             VAL VAL LYS GLU ALA LEU LEU LYS THR
CTA ATC CAT TAT TAA TGT TTT TTA TAT TTT GTAG/GTG GTT AAA GAA GCA CTG CTG AAA ACA

                                 1230                                      1260
ILE LYS GLU ALA VAL GLY GLY ASN TRP SER ASP GLU LEU SER SER ALA TRP GLU VAL ALA
ATA AAG GAA GCT GTT GGC GGC AAT TGG AGT GAC GAA TTG AGC AGT GCT TGG GAA GTA GCC

                                 1290
TYR ASP GLU LEU ALA ALA ALA ILE LYS LYS ALA
TAT GAT GAA TTG GCA GCA GCA ATT AAA AAG GCA TAA
```

The amino acid sequence of the Lbc_1 protein is indicated above the coding sequence,

Lbc$_2$ with the sequence:

```
                                                                      GLY
                                                                    G/GGT
                                                                      180

ALA PHE THR GLU LYS GLN GLU ALA LEU VAL SER SER SER PHE GLU ALA PHE LYS ALA ASN
GCT TTC ACT GAG AAG CAA GAG GCT TTG GTG AGT AGC TCA TTC GAA GCA TTC AAG GCA AAC
                                    210                                     240

ILE PRO GLN TYR SER VAL VAL PHE TYR THR SER
ATT CCT CAA TAC AGC GTT GTG TTC TAC ACT TC/GTA AGT TTT CTC TTA AAG CAT GTA TCT
                                    270                                     300

TTC ATT CTC TGT TTT TCC TTT CGA CAT TTT TTG TGT TTG AAA AGA GAT AGT GTC AAT GTG
                                    330                                     360

                                           ILE LEU GLU LYS ALA PRO ALA ALA LYS
AGT GGG TAT TTT TTT TTA TTA AAA ATT AAC AG/G ATA CTG GAG AAA GCA CCC GCA GCA AAG
                                    390                                     420

ASP LEU PHE SER PHE LEU SER ASN GLY VAL ASP PRO SER ASN PRO LYS LEU THR GLY HIS
GAC TTG TTC TCG TTT CTA TCT AAT GGA GTA GAT CCT AGT AAT CCT AAG CTC ACG GGC CAT
                                    450                                     480

ALA GLU LYS LEU PHE GLY LEU
GCT GAA AAG CTT TTT GGA TTG/GTA AGT ATC ATC CAA CTA AAA TTA TAG CTA TTT TAT GTG
                                        510                                 540

ATT AAT TTT AAG ATT AAA CAT GTA TTT AAC ACT CTT AAA CAT GTA TTT AAC ACT CTT AAG
                                    570                                     600

ATT AAA CAT GTA TTT AAC TAA AAC ATG TAT TTG CTG ATT ATT TTT TTT TTA TAA TTA TCT
                                    630                                     660

                                       VAL ARG ASP SER ALA GLY GLN LEU LYS ALA
TGT CAC ATA TTA TAT ATT TTT TGA ATT GTA G/GTG CGT GAC TCA GCT GGT CAA CTT AAA GCA
                                    690                                       720

ASN GLY THR VAL VAL ALA ASP ALA ALA LEU GLY SER ILE HIS ALA GLN LYS ALA ILE THR
AAT GGA ACA GTA GTG GCT GAT GCC GCA CTT GGT TCT ATC CAT GCC CAA AAA GCA ATC ACT
                                750                                         780

ASP PRO GLN PHE VAL
GAT CCT CAG TTC GTG/GT ATG ATA AAT AAT AAA ATG TTA CAA TAA ATG CAC ATA TAC TTA
                                    810                                     840

AAT TTT ACA TGG TGC AGT GTT ATG ATC ATC ATT TTT GTT TAG TAA TGA ATT TAC TTA AAA
                                    870                                     900

TCT TAA ATT ATG TAC TTT TTG AAA GTT TTA TAT GGA ATT TTA ATT ATA GGG AAA AAT GTA
                                    930                                     960

                                           VAL VAL LYS GLU ALA LEU LEU LYS THR
AGA GCT AAT CCA TTA GTG ATG TTT TGT CTG TAG/GTG GTT AAA GAA GCA CTG CTG AAA ACA
                                    990                                    1020

ILE LYS GLU ALA VAL GLY ASP LYS TRP SER ASP GLU LEU SER SER ALA TRP GLU VAL ALA
ATA AAG GAG GCA GTT GGG GAC AAA TGG AGT GAT GAA TTG AGC AGT GCT TGG GAA GTA GCC
                                1050                                       1080

TYR ASP GLU LEU ALA ALA ALA ILE LYS LYS ALA PHE
TAT GAT GAA TTG GCA GCA GCT ATT AAG AAG GCA TTT TAC
                                1110
```

The amino acid sequence of the Lbc_2 protein is indicated above the coding sequence,

and Lbc_3 with the sequence:

```
                                                                GLY ALA PHE THR ASP
                                                              G/GGT GCT TTC ACT GAT
                                                                                120

LYS GLN GLU ALA LEU VAL SER SER SER PHE GLU ALA PHE LYS THR ASN ILE PRO GLN TYR
AAG CAA GAG GCT TTG GTG AGT AGC TCA TTT GAA GCA TTC AAG ACA AAC ATT CCT CAA TAC
                                    150                                     180

SER VAL VAL PHE TYR THR SER
AGT GTT GTG TTC TAC ACC TC/GTA AGT ATT CTA TCT AAA TTA TGT GTC TTA TTG TAT GTT
                                    210                                     240

TAA CTT TCG TGG TTT GTT GTG TTT GAA AAA AAG ATA TAT ATT GTT AAT GTG AGT GGT TTT
                                    270                                     300

                              ILE LEU GLU LYS ALA PRO VAL ALA LYS ASP LEU PHE SER
GGT TTG ACT AAA AAT GAA TAG/G ATA CTG GAG AAA GCA CCT GTA GCA AAG GAC TTG TTC TCA
                                    330                                         360

PHE LEU ALA ASN GLY VAL ASP PRO THR ASN PRO LYS LEU THR GLY HIS ALA GLU LYS LEU
TTT CTA GCT AAT GGA GTA GAC CCC ACT AAT CCT AAG CTC ACG GGC CAT GCT GAA AAA CTT
                                    390                                     420

PHE GLY LEU
TTT GGA TTG/GT AAG TAT CCA GCC TAC TAA AAT TAA AAT CCT ATT AGT ATT TTT TAT TAT
                                    450                                     480

                                                                VAL ARG ASP SER
TTT TCT TCC ATG ATT GTC TTG TCA CAT ATT ATA TAT TTT TTG AAT TAT AG/GTA CGT GAT TCA
                                    510                                         540

ALA GLY GLN LEU LYS ALA SER GLY THR VAL VAL ILE ASP ALA ALA LEU GLY SER ILE HIS
GCT GGT CAA CTT AAA GCA AGT GGA ACA GTG GTG ATT GAT GCC GCA CTT GGT TCT ATC CAT
                                    570                                     600

ALA GLN LYS ALA ILE THR ASP PRO GLN PHE VAL
GCC CAA AAA GCA ATC ACT GAT CCT CAA TTT GTG/G TAT GAT AAA TAA TGA AAA GCT ACA
                                    630                                     660

ATA AAT GCA CAA ATA CTT AAT TTT ACA TAG TGC AGT GCT ATA TGA TCA TCA CTT TTG CTT
                                    690                                     720

AGT AAT GAA TTT ACT TTT TTT TTT TAC AGA AGT AAT GGA TTT ACT TAA AAT CTT AAA TTA
                                    750                                     780

TGT ACT TCT TTA AAG AGT TTT GTA TGG AAT TTT AAT TAT AGG AAA AAT GTA AGA GCT AAA
                                    810                                     840

                             VAL VAL LYS GLU ALA LEU LEU LYS THR ILE LYS GLU ALA
CCA TTG CTG ATG ATT TCG AAG/GTG GTT AAA GAA GCA CTG CTG AAA ACA ATA AAG GAG GCA
                                    870                                     900

VAL GLY ASP LYS TRP SER ASP GLU LEU SER SER ALA TRP GLU VAL ALA TYR ASP GLU LEU
GTT GGG GAC AAA TGG AGT GAC GAG TTG AGC AGT GCT TGG GAA GTA GCC TAT GAT GAA TTG
                                    930                                     960

ALA ALA ALA ILE LYS LYS ALA PHE
GCA GCA GCT ATT AAG AAG GCA TTT TAG
```

The amino acid sequence of the Lbc_3 protein is indicated above the coding sequence.

The present invention furthermore deals with a novel DNA fragment comprising an inducible plant promoter to be used when carrying out the method according to the invention, said DNA fragment being characterised by being identical with, derived from or comprising a 5' flanking region of root nodule-specific genes. Examples of such DNA fragments are DNA fragments being identical with, derived from or comprising a 5' flanking region of plant leghemoglobin genes. Preferred examples are according to the invention DNA fragments being identical with, derived from or comprising a 5' flanking region of the four soybean leghemoglobin genes, viz.:

Lba with the sequence:

```
GAGATACATT ATAATAATCT CTCTAGTGTC TATTTATTAT TTTATCTGGT
GATATATACC TTCTCGTATA CTGTTATTTT TTCAATCTTG TAGATTTACT
TCTTTTATTT TTATAAAAAA GACTTTATTT TTTTAAAAAA AATAAAGTGA
ATTTTGAAAA CATGCTCTTT GACAATTTTC TGTTTCCTTT TTCATCATTG
GGTTAAATCT CATAGTGCCT CTATTCAATA ATTTGGGCTC AATTTAATTA
GTAGAGTCTA CATAAAATTT ACCTTAATAG TAGAGAATAG AGAGTCTTGG
AAAGTTGGTT TTTCTCGAGG AAGAAAGGAA ATGTTAAAAA CTGTGATATT
TTTTTTTTGG ATTAATAGTT ATGTTTATAT GAAAACTGAA AATAAATAAA
CTAACCATAT TAAATTTAGA ACAACACTTC AATTATTTTT TTAATTTGAT
TAATTAAAAA ATTATTTGAT TAAATTTTTT AAAAGATCGT TGTTTCTTCT
TCATCATGCT GATTGACACC CTCCACAAGC CAAGAGAAAC ACATAAGCTT
TGGTTTTCTC ACTCTCCAAG CCCTCTATAT AAACAAATAT TGGAGTGAAG
TTGTTGCATA ACTTGCATCG AACAATTAAT AGAAATAACA GAAAATTAAA
AAAGAAATAT G,
```

Lbc_1 with the sequence

TTCTCTTAAT ACAATGGAGT TTTTGTTGAA CATACATACA TTTAAAAAAA
AATCTCTAGT GTCTATTTAC CCGGTGAGAA GCCTTCTCGT GTTTTACACA
CTTTAATATT ATTATATCCT CAACCCCACA AAAAAGAATA CTGTTATATC
TTTCCAAACC TGTAGATTTA TTTATTTATT TATTTATTTT TACAAAGGAG
ACTTCAGAAA AGTAATTACA TAAAGATAGT GAACATCATT TTATTTATTA
TAATAAACTT TAAAATCAAA CTTTTTTATA TTTTTTGTTA CCCTTTTCAT
TATTGGGTGA AATCTCATAG TGAAGCCATT AAATAATTTG GGCTCAAGTT
TTATTAGTAA AGTCTGCATG AAATTTAACT TAACAATAGA GAGAGTTTTC
GAAAGGGAGC GAATGTTAAA AAGTGTGATA TTATATTTTA TTTCGATTAA
TAATTATGTT TACATGAAAA CATACAAAAA AATACTTTTA AATTCAGAAT
AATACTTAAA ATATTTATTT GCTTAATTGA TTAACTGAAA ATTATTTGAT
TAGGATTTTG AAAAGATCAT TGGCTCTTCG TCATGCCGAT TGACACCCTC
CACAAGCCAA GAGAAACTTA AGTTGTAAAC TTTCTCACTC CAAGCCTTCT
ATATAAACAT GTATTGGATG TGAAGTTATT GCATAACTTG CATTGAACAA
TAGAAAATAA CAAAAAAAAG TAAAAAAGTA GAAAAGAAAT ATG,

Lbc_2 with the sequence:

TCGAGTTTTT ACTGAACATA CATTTATTAA AAAAAACTCT CTAGTGTCCA
TTTATTCGGC GAGAAGCCTT CTCGTGCTTT ACACACTTTA ATATTATTAT
ATCCCCACCC CCACCAAAAA AAAAAAAACT GTTATATCTT TCCAGTACAT
TTATTTCTTA TTTTTACAAA GGAAACTTCA CGAAAGTAAT TACAAAAAAG
ATAGTGAACA TCATTTTTTT AGTTAAGATG AATTTTAAAA TCACACTTTT
TTATATTTTT TTGTTACCCT TTTCATTATT GGGTGAAATC TCATAGTGAA
ACTATTAAAT AGTTTGGGCT CAAGTTTTAT TAGTAAAGTC TGCATGAAAT
TTAACTTAAT AATAGAGAGA GTTTTGGAAA GGTAACGAAT GTTAGAAAGT
GTGATATTAT TATAGTTTTA TTTAGATTAA TAATTATGTT TACATGAAAA
TTGACAATTT ATTTTTAAAA TTCAGAGTAA TACTTAAATT ACTTATTTAC
TTTAAGATTT TGAAAAGATC ATTTGGCTCT TCATCATGCC GATTGACACC
CTCCACAAGC CAAGAGAAAC TTAAGTTGTA ATTTTTCTAA CTCCAAGCCT
TCTATATAAA CACGTATTGG ATGTGAAGTT GTTGCATAAC TTGCATTGAA
CAATAGAAAT AACAACAAAG AAAATAAGTG AAAAAAGAAA TATG,

and Lbc_3 with the sequence:

TATGAAGATT AAAAAATACA CTCATATATA TGCCATAAGA ACCAACAAAA
GTACTATTTA AGAAAAGAAA AAAAAAACCT GCTACATAAT TTCCAATCTT
GTAGATTTAT TTCTTTTATT TTTATAAAGG AGAGTTAAAA AAATTACAAA
ATAAAAATAG TGAACATCGT CTAAGCATTT TTATATAAGA TGAATTTTAA
AAATATAATT TTTTTGTCTA AATCGTATGT ATCTTGTCTT AGAGCCATTT
TTGTTTAAAT TGGATAAGAT CACACTATAA AGTTCTTCCT CCGAGTTTGA
TATAAAAAAA ATTGTTTCCC TTTTGATTAT TGGATAAAAT CTCGTAGTGA
CATTATATTA AAAAAATTAG GGCTCAATTT TTATTAGTAT AGTTTGCATA
AATTTTAACT TAAAAATAGA GAAAATCTGG AAAAGGGACT GTTAAAAAGT
GTGATATTAG AAATTTGTCG GATATATTAA TATTTTATTT TATATGGAAA
CTAAAAAAAT ATATATTAAA ATTTTAAATT CAGAATAATA CTTAAATTAT
TTATTTACTG AAAATGAGTT GATTTAAGTT TTTGAAAAGA TGATTGTCTC
TTCACCATAC CAATTGATCA CCCTCCTCCA ACAAGCCAAG AGAGACATAA
GTTTTATTAG TTATTCTGAT CACTCTTCAA GCCTTCTATA TAAATAAGTA
TTGGATGTGA AGTTGTTGCA TAACTTGCAT TGAACAATTA ATAGAAATAA
CAGAAAAGTA GAAAAGAAAT ATG.

Another example of a preferred DNA fragment according to the invention is a DNA fragment which is identical with, derived from or comprises 5' flanking regions of the Lbc_3-5'-3'CAT gene with the sequence

TATGAAGATT AAAAAATACA CTCATATATA TGCCATAAGA ACCAACAAAA
GTACTATTTA AGAAAAGAAA AAAAAAACCT GCTACATAAT TTCCAATCTT
GTAGATTTAT TTCTTTTATT TTTATAAAGG AGAGTTAAAA AAATTACAAA
ATAAAAATAG TGAACATCGT CTAAGCATTT TTATATAAGA TGAATTTTAA
AAATATAATT TTTTTGTCTA AATCGTATGT ATCTTGTCTT AGAGCCATTT
TTGTTTAAAT TGGATAAGAT CACACTATAA AGTTCTTCCT CCGAGTTTGA
TATAAAAAAA ATTGTTTCCC TTTTGATTAT TGGATAAAAT CTCGTAGTGA
CATTATATTA AAAAAATTAG GGCTCAATTT TTATTAGTAT AGTTTGCATA
AATTTTAACT TAAAAATAGA GAAAATCTGG AAAAGGGACT GTTAAAAAGT
GTGATATTAG AAATTTGTCG GATATATTAA TATTTTATTT TATATGGAAA
CTAAAAAAAT ATATATTAAA ATTTTAAATT CAGAATAATA CTTAAATTAT
TTATTTACTG AAAATGAGTT GATTTAAGTT TTTGAAAAGA TGATTGTCTC
TTCACCATAC CAATTGATCA CCCTCCTCCA ACAAGCCAAG AGAGACATAA
GTTTTATTAG TTATTCTGAT CACTCTTCAA GCCTTCTATA TAAATAAGTA
TTGGATGTGA AGTTGTTGCA TAACTTGCAT TGAACAATTA ATAGAAATAA
CAGAAAAGTA GAATTCTAAA ATG

Still an ther example of such a DNA fragment ac-

cording to the invention is a DNA fragment which is identical with, derived from or comprises 5' flanking regions of the N23 gene with the sequence

```
           10        20        30        40        50        60        70
GAATTCGAGCTCGCCCGGGGATCGATCCTCTAGAGTCGACCTGCAGCCCAAGCTTGGATCAATCAATTAA
 EcoRI                             SalI

           80        90       100       110       120       130       140
TTCTATTGAGACACGATTTGAACAATTTTTACATTATGAGACTATTTTTGGTTTTTTATTTGATCCAAAA

          150       160       170       180       190       200       210
AAATTTAAAGCTTTAGATGATGATGAATTGAANNAATATTGTATTAATNNTGAAAAGTTNNNNNGGTTTA

          220       230       240       250       260       270       280
ATGAATGCTATGATATTGATGGTCTTGATNTATTNNCAGAATTGAAAGTATTAAGAGAAGTGTTAAGAAA

          290       300       310       320       330       340       350
AGAAGTTAGCACACCAATAGAAGTATTGAGTTATATTAAAACTTTAGATTCTTTTCAAATGTTTACATTG

          360       370       380       390       400       410       420
CATATAGAATTTTATTGACAATCCTTATAACAGTTGCTACTGTTGAAAGACGTTCTTCAAAATTAAAATT

          430       440       450       460       470       480       490
ACTTAAATCATATCTAAAATCAACAATGTTACAAGATAGATTGAATGAGTTAGTTATTTTATCTATTGAA

          500       510       520       530       540       550       560
AGTAAAGTGTTAGAATTGTTTGATTATAAAACTCTGATAAATGATTTTGCAGTTAAAAAAACTAGAAGAT

          570       580       590       600       610       620       630
TAATATAAAAATTGATATTTTATATAATATATTAAGTCTCTTTAAAATTCTTGTAAAAAAAGACATTTTT

          640       650       660       670       680       690       700
AAATAATAAAATAAAGCAACTCTTAATTTTAATGAAACATCCCTTTGTTAAACCGAATCTTCCATAATGT

          710       720       730       740       750       760       770
AAAAATTAATGCTTGATGGAAGTTTTTAATTTGTTCTACTCAATACTCAAAGGGTTGTAAATATTTTTTT

          780       790       800       810       820       830       840
TATCATTTATATGTTGTAAATATGAATGCACTAGTAATTAGTTTAATGATAAAATATATTCTACAGATAT

          850       860       870       880       890       900       910
ATTTCTGTCTCTTGGCAACTCGTGAGAATTGAATATATTATAAAGATGAAAGGTCGTTACAATTTTTTTT

          920       930       940       950       960       970       980
AGAATAAATATTTATATACAATTCCTAGATTTTGTTATAAAATTCACATATTGTATGAGTATAAATACAT

          990      1000      1010      1020      1030      1040      1050
GAGCACACACCAAACTAGTCTCAAATTAAGTAAGGTGCTAATTATTAGCGGCTAGCTAAGTAACCAAGTA
                                                         DdeI

ATTAATG
```

The invention relates furthermore to any plasmid to be used when carrying out the method according to the invention and characterised by comprising a DNA fragment containing an inducible plant promoter as herein defined. Particular examples of suitable plasmids according to the invention are pAR11, pAR29, pAR30, and N23-CAT, cf. Examples 3, 4, and 11. These plasmids allow recombination into the A. rhizogenes T DNA region.

The invention relates furthermore to any Agrobacterium strain to be used in connection with the invention and characterised by comprising a DNA fragment comprising an inducible plant promoter of root nodule-specific genes built into the T DNA region and therefore capable of transforming the inducible promoter into plants. Particular examples of bacterium strains according to the invention are the A. rhizogenes strains AR1127 carrying pAR29, AR1134 carrying pAR30, AR1000 carrying pAR11, and AR204-N23-CAT carrying N23-CAT.

It is obvious that the patent protection of the present invention is not limited by the embodiments stated above.

Thus the invention employs not exclusively 5' flanking regions of soybean leghemoglobin genes. It is well-known that the leghemoglobin genes of all leguminous plants have the same function, cf. Appleby (1974) in The Biology of Nitrogen Fixation, Quispel. A. Ed. North-Holland Publishing Company, Amsterdam, Oxford, pages 499-554, and concerning the kidney bean PvLb1 gene it has furthermore been

proved that a high degree of homogoly exists with the sequences of the soybean Lbc_3 gene. It is also known that the expression of other root nodule-specific genes is regulated in a similar manner like the leghemoglobin genes. The invention includes thus the use of 5' flanking regions of leghemoglobin genes or other root nodule-specific genes of all plants in case the use of such DNA fragments makes the expression of a desired gene product the subject matter of the regulation characterised by the present invention.

The present invention allows also the use of such fragments of any origin which under natural conditions exert or mediate the regulation characterised by the present invention. The latter applies especially to such fragments which can be isolated from DNA fragments from gene libraries or genomes through hybridization with labelled sequences of 5' flanking regions of soybean leghemoglobin genes.

It is well-known that it is possible to alter nucleotide sequences of non-important sub-regions of 5' flanking regions without causing an alteration of the promoter activity and the regulation. It is also well-known that an alteration of sequences of important subregions of 5' flanking regions renders it possible to alter the binding affinities between nucleotide sequences and the factors or effector substances necessary or responsible for the transcription initation and the translation initiation and consequently to improve the promoter activity and/or the regulation. The present invention includes, f course, also the use of DNA fragments

containing such altered sequences of 5'flanking regions, and in particular DNA fragments can be mentioned which have been produced by recombining sequences of 5' flanking regions of any gene with 5' flanking regions of root nodule-specific genes provided the use of such DNA fragments subjects the expression of a desired gene product to the regulation characterised by the present invention.

It should be noted that the transformation of micro-organisms is carried out in a manner known per se, cf. e.g. Maniatis et al., (1982), Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory.

The transformation of plant cells, i.e. introduction of plasmid DNA into plant cells, is also carried out in a manner known per se, cf. Zambryski et al., (1983), EMBO J. 2, 2143-2150.

Cleavage with restriction endonucleases and digestion with other DNA modifying enzymes are well-known techniques and are carried out as recommended by the suppliers.

The _Agrobacterium rhizogenes_ 15834 rif^R was used as a typical representative of A. rhizogenes: see White et al., I.Bact., Vol. 141 (1980), 1134-1141.

Example 1

Sequence determination of 5' flanking regions of soybean leghemoglobin genes

From a soybean gene library the four soybean leg-

hemoglobin genes Lba, Lbc_1, Lbc_2, and Lbc_3 are provided as described by Jensen, E.Ø. et al., Nature Vol. 291, No. 3817, 677-679 (1981). The genetically stable in-bred invariable soybean species "Glycine max.var.Evans" was used as a starting material for the isolation of the DNA used for the construction of said gene library. The 5' flanking regions of the four soybean leghemoglobin genes are isolated, as described by Jensen, E.Ø., Ph D Thesis, Institut for Molekylær Biologi, Århus Universitet (1985), and the DNA sequences determined by the use of the dideoxy method as described by Sanger, F., J. Mol. Bio. 143, 161-178 (1980) and indicated in the sequence scheme.

Example 2

Construction of Lbc_3-5'-3'-CAT

The construction has been carried out in a sequence of process steps as described below:

a) Sub-cloning the Lbc_3 gene

The Lbc_3 gene was isolated on a 12Kb EcoRI restriction fragment from a soybean DNA library, which has been described by Wiborg et al., in Nucl. Acids Res. (1982) 10, 3487. A section of the fragment is shown at the top of the attached Scheme 2. This fragment was digested by the enzymes stated and then ligated to pBR322 as indicated at the Scheme. The resulting plasmids Lbc_3HH and Lbc_3HX were subsequently digested by PvuII and religated, which resulted in two plasmids called pLpHH and pLpHX.

b) Sub-cloning 5'flanking sequences from the Lbc_3 gene

For this purpose pLpHH was used as shown in the attached Scheme 3. This plasmid was opened by means of PvuII and treated with exonuclease Bal31. The reaction was stopped at various times and the shortened plasmids were ligated into fragments from pBR322. These fragments had been treated in advance as shown in Scheme 3, in such a manner that in one end they had a DNA sequence TTC ---
AAG ---.

After the ligation a digestion with EcoRI took place, and the fragments containing 5' flanking sequences were ligated into EcoRI digested pBR322. These plasmids were transformed into E. coli K803, and the plasmids in the transformants were tested by sequence analysis. A plasmid, p213 5'Lb, isolated from one of the transformants, contained a 5' flanking sequence terminating 7 bp before the Lb ATG start codon in such a manner that the sequence is as follows:

2Kb
-5' flanking --- AAAGTAGAATTC
Lbc_3 sequence

E.coli K803 is a typical representative of the E. coli K12 recipient strains.

c) Sub-cloning 3' flanking region of the Lbc_3 gene

For this purpose pLpHX was used which was digested by XhoII. The ends were partially filled out and excess single-stranded DNA was removed with S1 nuclease, as shown in the attached Scheme 4. The fragment shown was ligated into pBR322 which had been pretreated as shown in the Scheme. The construction was transformed into E. coli K803. One of the transformants contained a plasmid called Xho2a-3'Lb. As the XhoII recognition sequence is positioned immediately after the Lb stop codon, cf. Scheme 2, the plasmid contained about 900 bp of the 3' flanking region, and the sequence started with GAATTCTACAA---.

The construction of Lb promoter cassette

An EcoRI/SphI fragment from Xho2a-3'Lb was mixed with a BamHI/EcoRI fragment from p213-5'Lb. These two fragments were ligated via the BamHI/SphI cleavage sites into a pBR322 derivative where the EcoRI recognition sequence had been removed, cf. Scheme 4. The ligated plasmids were transformed into E. coli K803. A plasmid in one of the transformants contained the correct fragments, and it was called pEJLb 5'-3'-1.

Construction of the Lbc_3 5'3'-CAT gene

The CAT gene of pBR322 was isolated on several smaller restriction fragments, as shown in the attached Scheme 5. The 5' coding region was isolated as an AluI fragment which was subsequently ligated into pBR322, treated as stated in the Scheme. This

was transformed into E. coli K803. Several transformants contained the correct plasmid. One was taken out and called Alu11. The 3' coding region was isolated on a TaqI fragment. This fragment was treated with exonuclease Bal31, whereafter EcoRI linkers were added. Then followed a digestion with EcoRI and a ligation to EcoRI digested pBR322. The latter was transformed into E. coli K803 and the transformants were analysed. A plasmid, Taq 12, contained the 3' coding region of the CAT gene plus 23 bp 3' flanking sequences subsequently terminating in the following sequence CCCCGAATTC. Subsequently the following fragments were ligated together to EcoRI digested pEJLb5'-3'-1: EcoRI/PvuII fragment from AluI, PvuII/Ddel fragment from pBR322 and DdeI/EcoRI fragment from Taq 12. This ligation mixture was transformed into E. coli K803. Several transformants contained the correct plasmid. One was taken out and was called pEJLb 5'-3' CAT 15.

Example 3

a.

Cloning and integration of the soybean Lbc_3-5'-3'-CAT gene.

Two EcoRI fragments (No. 36 and No. 40) of the T_L-DNA region of A. rhizogenes 15834 pRi plasmid was used as "integration sites". Thus the Lbc_3-5'-3-CAT gene was subcloned (as 3,6 Kb BamHI/SalI fragment) into two vectors pAR1 and pAR22 carrying the ab ve EcoRI fragments. The resulting plasmids pAR29

and pAR30 were separately mobilized into _A. rhizogenes_ 15834 rif^R using a plasmid helper system; see E. van Haute et al. (1983), EMBO J. 3, 411-417. Neither pAR29 nor pAR30 can replicate in Agrobacterium. Therefore the selection by means of rifampicin 100 µg/ml and the plasmid markers spectinomycine 100 µg/ml, streptomycine 100 µg/ml or kanamycine 300 µg/ml will select _A. rhizogenes_ bacteria having integrated the plasmids via homologous recombination through the EcoRI fragments 36 or 40. The structure of the resulting T_L-DNA regions - transferred to the transformed plant lines L5-9 and L6-23 - has been indicated at the bottom of the attached Scheme 6. In this Scheme is furthermore for the L6-23 line shown the EcoRI and HindIII fragments carrying the Lbc_3-5'-3'-CAT gene and therefore hybridizing to radioactively labelled Lbc_3-5'-3'-CAT DNA used as a probe, cf. Example 4_a_.

_b._

Cloning and integration of the soybean Lbc_3 gene.

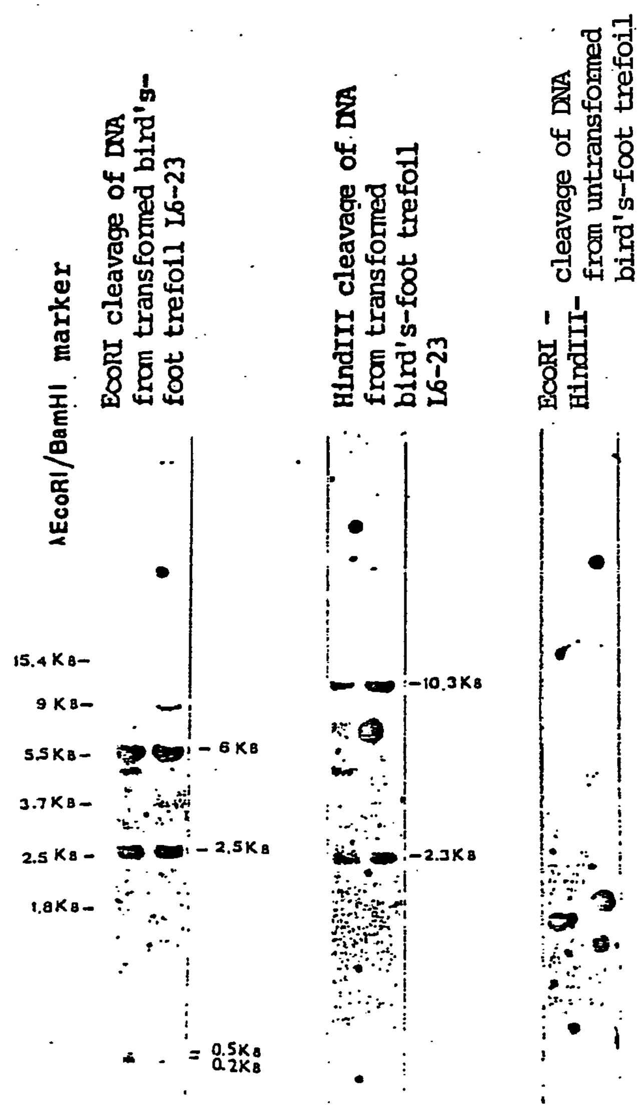
The EcoRI fragment No. 40 has here been used as "integration site". The Lbc_3 gene was therefore sub-cloned (as a 3,6 Kb BamHI fragment into the pAR1 vector and transferred into the T_L-DNA region as stated in a. The structure of the T_L-DNA region, transferred to the transformed plant line L8-35, has been shown at the bottom of the attached Scheme 7. This Scheme furthermore shows the EcoRI and HindIII fragments carrying the Lbc_3 gene and there-

fore hybridizing with radioactively labelled Lbc_3 DNA used as a probe, cf. Example 4<u>b</u>.

Example 4.

a.

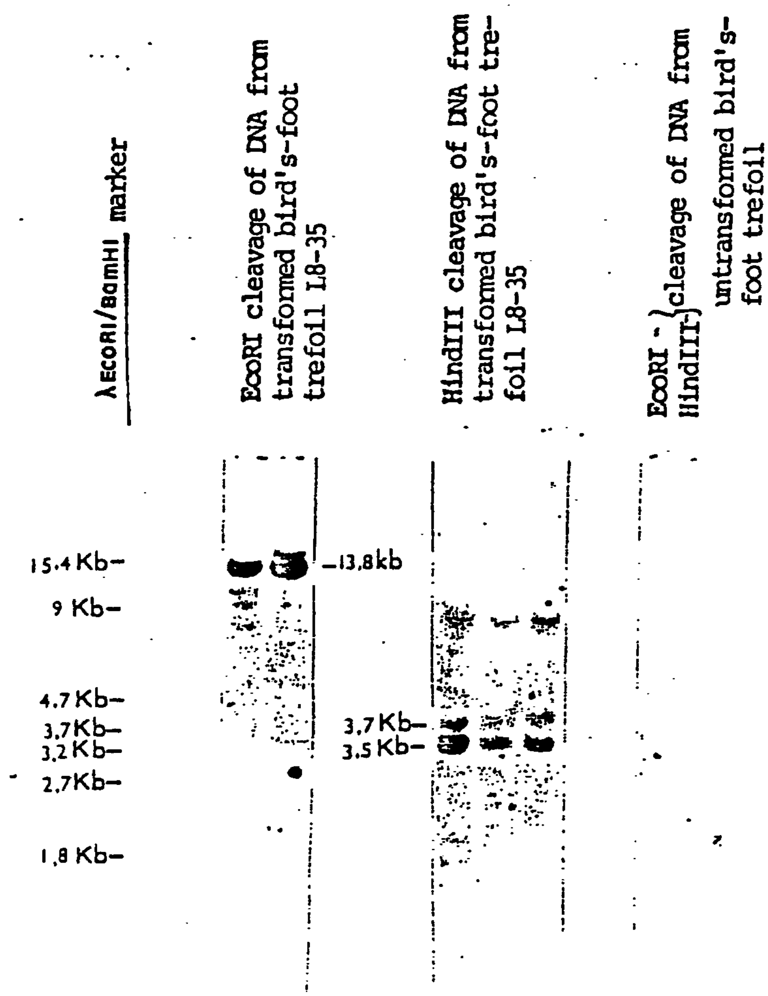
Demonstration of the soybean Lbc_3-5'-3'-CAT gene in transformed plants of bird's-foot trefoil.



DNA extracted from transformed lines (L6-23) or untransformed control plants and cleaved by the restriction enzymes EcoRI and HindIII was analyzed by Southern-hybridization. Radioactively labelled Lbc_3-5'-3'-CAT gene was used as a probe for demonstrating corresponding sequences in the transformed lines. The bands marked with numbers correspond to restriction fragments constituting parts of the Lbc_3-5'-3'-CAT gene as stated in the restriction map (Scheme 6) of Example 3a.

b.

Demonstration of the soybean Lbc_3 gene of transformed plants of bird's-foot trefoil.

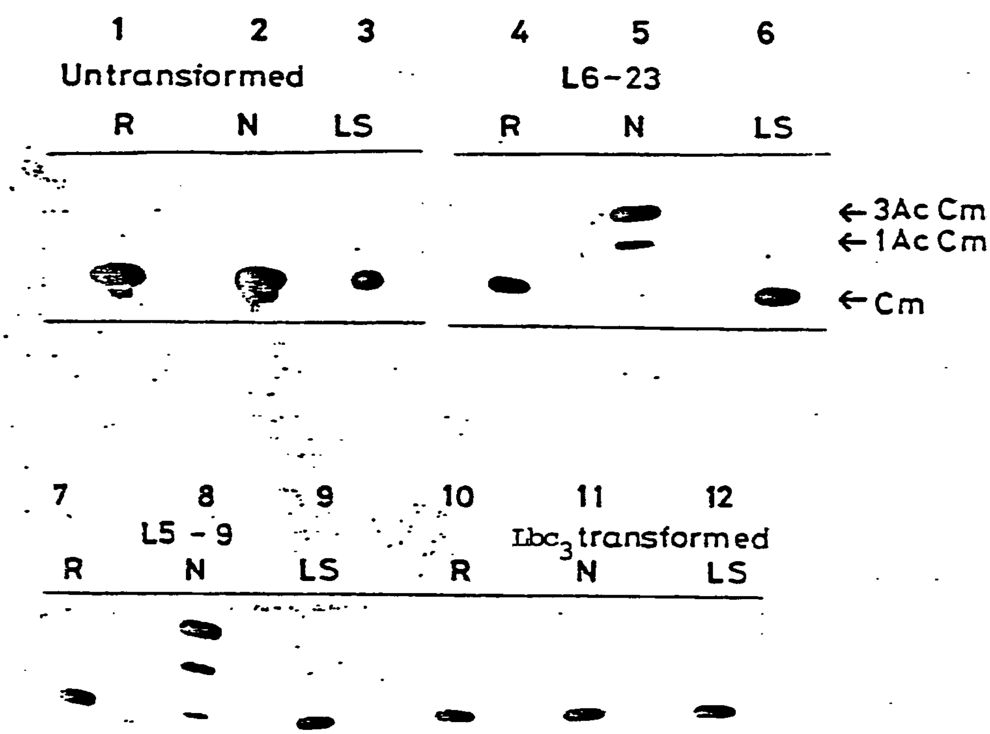


DNA extracted from transformed lines (L8-35) or untransformed control plants and cleaved by the restriction enzymes EcoRI and HindIII was analyzed by Southern-hybridization. Radioactive Lbc_3 gene was used as a probe for detecting corresponding sequences in the transformed lines. The bands marked with numbers correspond to restriction fragments constituting parts of the Lbc_3 gene as stated in the restriction map (Scheme 7) f Example 3b.

Example 5

a.

Expression of the Lbc_3-5'-3'-CAT gene in various tissues of bird's-foot trefoil.



The activity of the chloroamphenicol acetyl transferase (CAT) enzyme is measured as the amount of acetylated chloroamphenicol (AcCm) produced from ^{14}C-chloroamphenicol. In (a) the acetylated forms 1AcCm and 3AcCm appear, which have been separated from Cm through thin-layer chromatography in chloroform/methanol (95:5). The columns 1-3 show that no CAT activity occurs in root (R), nodule (N), as well as leaves + stem (LS) of untransformed plants of bird's-foot trefoil. The columns 4-6 and 7-9 show the CAT activity in corresponding tissues of Lbc_3-5'-3'-CAT transformed L6-23 and L5-9 plants. The conversion of chloroamphenicol in columns 5 and 8 shows the organ-specific expression of the Lbc_3-5'-3'-CAT gene in root nodules. The columns 10-12 show the lack of CAT activity in plants transformed with the Lbc_3 gene.

b.

Table

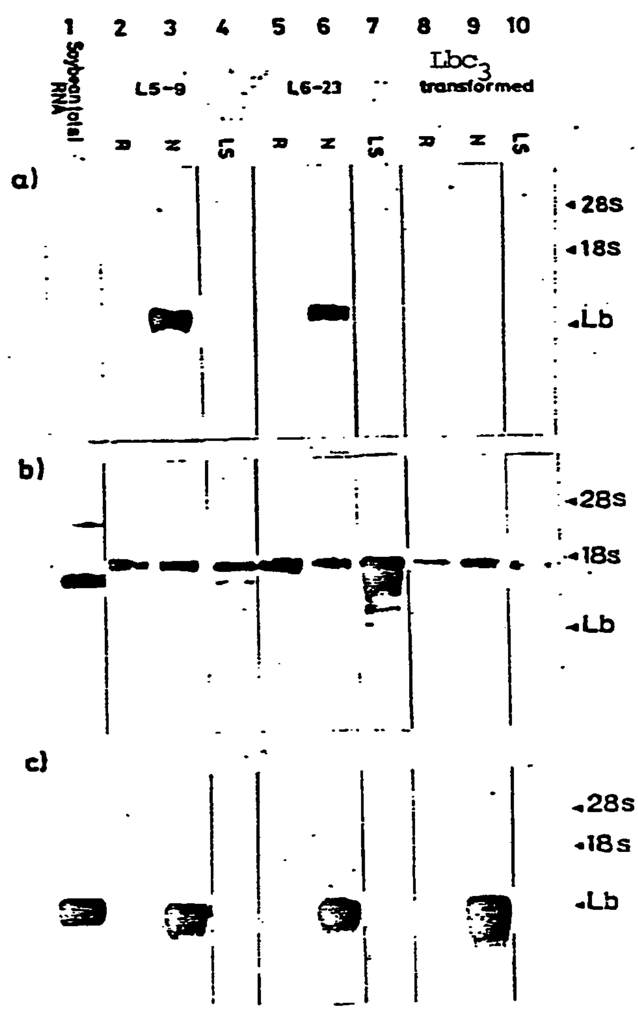
	L6-23 CAT activity	L5-9 CAT activity
Root	0	0
Nodule	68830 cpm/μg protein·h	154,000 cpm/μg protein·h
Leaves + Stem	0	0

In the Table (b) the CAT activity in Lbc_3-5'-3'-CAT transformed L5-9 and L6-23 plants has been stated as the amount of ^{14}C-chloroamphenicol converted into acetylated derivatives. The amount of radioactivity in the acetylated derivatives has been

counted by liquid scintillation and stated in cpm/μg protein · hour.

Example 6

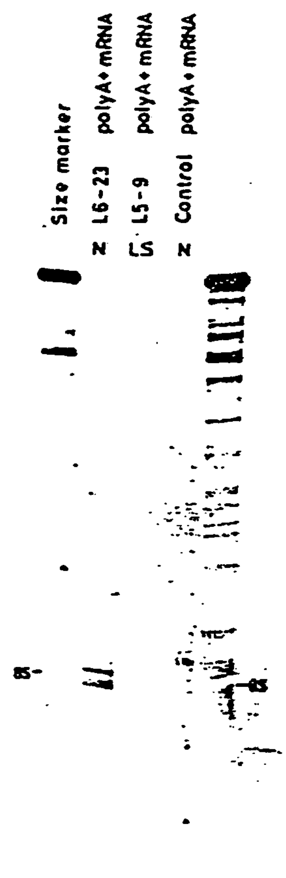
Transcription test (Northern analysis) on tissues of Lbc_3-5'-3'-CAT transformed and Lbc_3 transformed Lotus plant lines.



5 μg of total RNA extracted from root (R), nodule (N) or leaves + stem (LS) and separated in formaldehyde agarose gels were transferred onto nitrocellulose. Column 1 contains 5 μg of total RNA from 20-day-old soybean nodules as control plants. The columns 2-4 and 5-7 contain total RNA from root, nodule or leaves + stem, respectively, of the Lbc_3-5'-3'-CAT transformed lines L5-9 and L6-23. The columns 8-10 contain RNA from corresponding tissues of bird's-foot trefoil transformed by means of A. rhizogenes carrying the Lbc_3 gene in the T_L-DNA. In (a) radioactive DNA of the CAT coding sequence has been used as a probe for hybridization. The organ-specific transcription of the Lbc_3-5'-3'-CAT gene in root nodules from the L5-9 and L6-23 lines appears from columns 3 and 6. In (b) the transcript for the constitutive ubiquitine gene(s) is visualized using a cDNA probe for the human ubiquitine gene for the hybridization. In (c) the nodule-specific transcription of bird's-foot trefoil own leghemoglobin genes is shown. A cDNA probe of the Lba gene of soybean has been used for this hybridization.

Example 7

Determination of the transcription initiation site (CAP site) of the Lbc_3 promoter of soybean in transformed root nodules of bird's-foot trefoil.



—TATAAATAAGTATTTGGATGTGAAGTTGTTGCATAACT—/ /—AAATGGAG

83

The position of the "CAP site" was determined on the nucleotide level by means of primer extension. A synthetic oligonucleotide 5'CAACGGTGGTATATCCAGTG3' complementary to the nucleotides 15-34 in the coding sequence of the CAT gene was used as primer for the enzyme reverse transcriptase. As a result single-stranded cDNA was formed the length of which corresponds to the distance between the 5' end of the primer and the 5' end of the primed mRNA. A 83 nucleotide cDNA strand would be expected according to the knowledge of the transcription initiation site of soybean Lbc_3 gene. Columns 2, 3, and 4 from left to right show the produced DNA strands when the primer extension has been operated on $polyA^+$-purified mRNA from transformed root nodules of bird's-foot trefoil, transformed leaves + stem of bird's-foot trefoil, and untransformed root nodules of bird's-foot trefoil, respectively. The 85, 86, 87, 88, and 90 nucleotides long cDNA strand shown in column 2 proved correctly Lbc_3 promoter function in bird's-foot trefoil. The CAP sites corresponding to the cDNA sequences generated are indicated with asterisks (*) on the partial sequence of the Lbc_3 5'3'-CAT region given. In the sequence the TATA box of the Lbc_3 promoter and the corresponding translation initiation codon of the CAT coding sequence are underlined.

Example 8

Demonstration of the correct developmental control of the Lbc_3-5'-3'-CAT gene in transformed plants of bird's-foot trefoil (L6-23).

	Stage 1: No visible nodules	Stage 2: Emerging nodules	Stage 3: Distinct white nodules	Stage 4: Small pink nodules	Stage 5: Later stages of maturity
CAT activity in cpm/μg protein·hour	0	0	32.6	342.3	1255*
Nitrogenase activity nmol ethylene/μg protein · hour	0	O	0	O.5	2.7

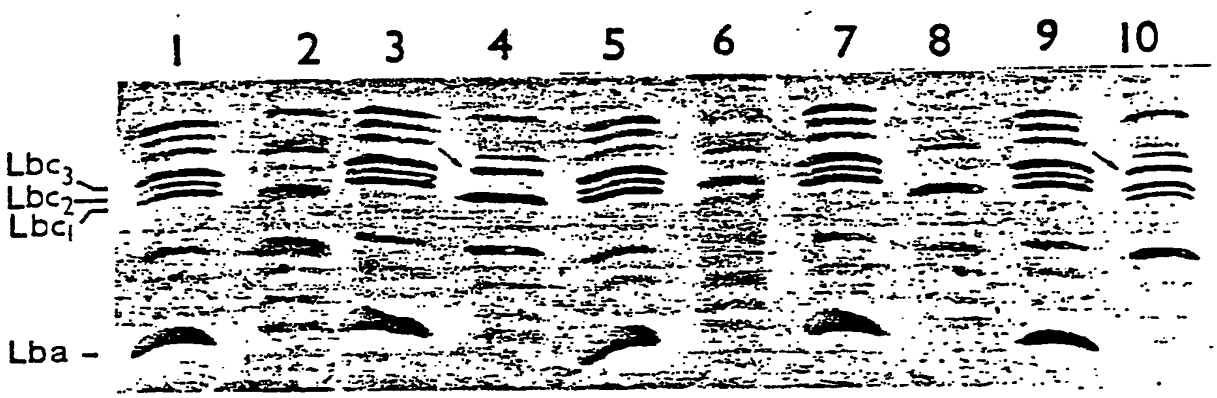
* Substrate limited reaction; actual activity about 68000 cpm/μg protein · hour.

Chloroamphenicol acetyl transferase and nitrogenase activity were measured on cut off pieces of root with nodules at the different developmental stages indicated. The CAT activity can be detected in the white distinct nodules whereas the nitrogenase activity did not appear until the small pink nodules have developed. The latter development corresponds to the development known from soybean control plants and described by Marcker et al. EMBO J. 1984, 3, 1691-95. The CAT activity was determined as in Example 5. The nitrog nase activity was measured

as acetylene reduction capacity of the nodules followed by gaschromatographic determination of ethylene.

Example 9

Demonstration of Lbc_3 protein in bird's-foot trefoil plants transformed with the soybean Lbc_3 gene.



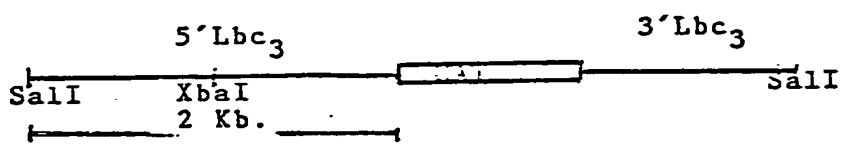
Proteins extracted from root nodules of Lbc_3 transformed (L8-35), Lbc_3-5'-3'-CAT transformed and nontransformed plants were separated by isolectric focussing at a pH gradient of 4 to 5. The columns 1, 3, 5, 7, and 9 show Lbc_1, Lbc_2, Lbc_3, and Lba proteins synthesized in soybean control root nodules. Column 2 shows proteins from root nodules of Lbc_3-5'-3'-CAT transformed L6-23-bird's-foot trefoil plants, whereas the columns 6 and 8 show proteins from nontransformed plants. The columns 4 and 10 show soybean Lbc_3 protein synthesized in root nod-

ules of bird's-foot trefoil plants (L8-35) transformed with the Lbc_3 gene. The Lbc_3 protein band is indicated by an arrow.

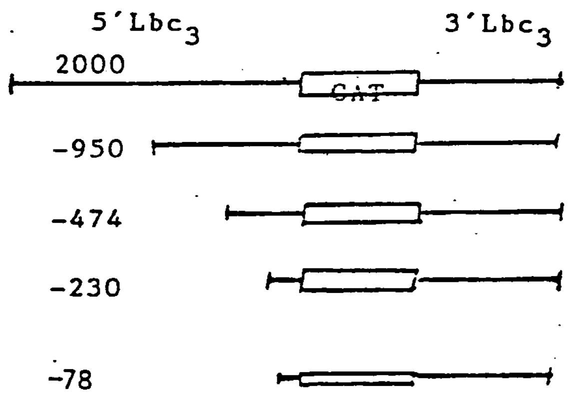
Example 10

Expression of the Lbc_3-5'-3'-CAT gene requires the 5' Lbc_3 promoter region.

The Lbc_3-5'-3'-CAT gene construction carries a 2 Kb 5' Lbc_3 promoter region. Stepwise removal of sequences from the 5' end of this region demonstrated that this promoter region is required for the characteristic expression of the Lbc_3-5'3'-CAT gene.



The Lbc_3-5'-3'-CAT gene construction was opened in the unique XbaI site shown above, and digested with the exonuclease Bal31. A SalI linker fragment was ligated onto the blunt ends generated and the shortened SalI fragments carrying the Lbc_3-5'-3'-CAT gene were transferred into L.corniculatus. The effect of removing promoter sequences was measured as CAT activity. End points of the deleted 5' region are given as the distance from the CAP site in nucleotides.



	CAT activity Cpm/µg protein/hrs.		
	Root	Nodule	Leaf
2000	0	80000	0
-950	0	10000	0
-474	0	3000	0
-230	0	3000	0
-78	0	0	0

The drastically reduced level of CAT activity expressed from the Lbc_3 promoter deleted to nucleotide -230 and the zero activity from the promoter deleted to nucleotide -78 demonstrates that the Lbc_3 promoter region is required for the root nodule specific expression of the Lbc_3-5'-3'-CAT gene.

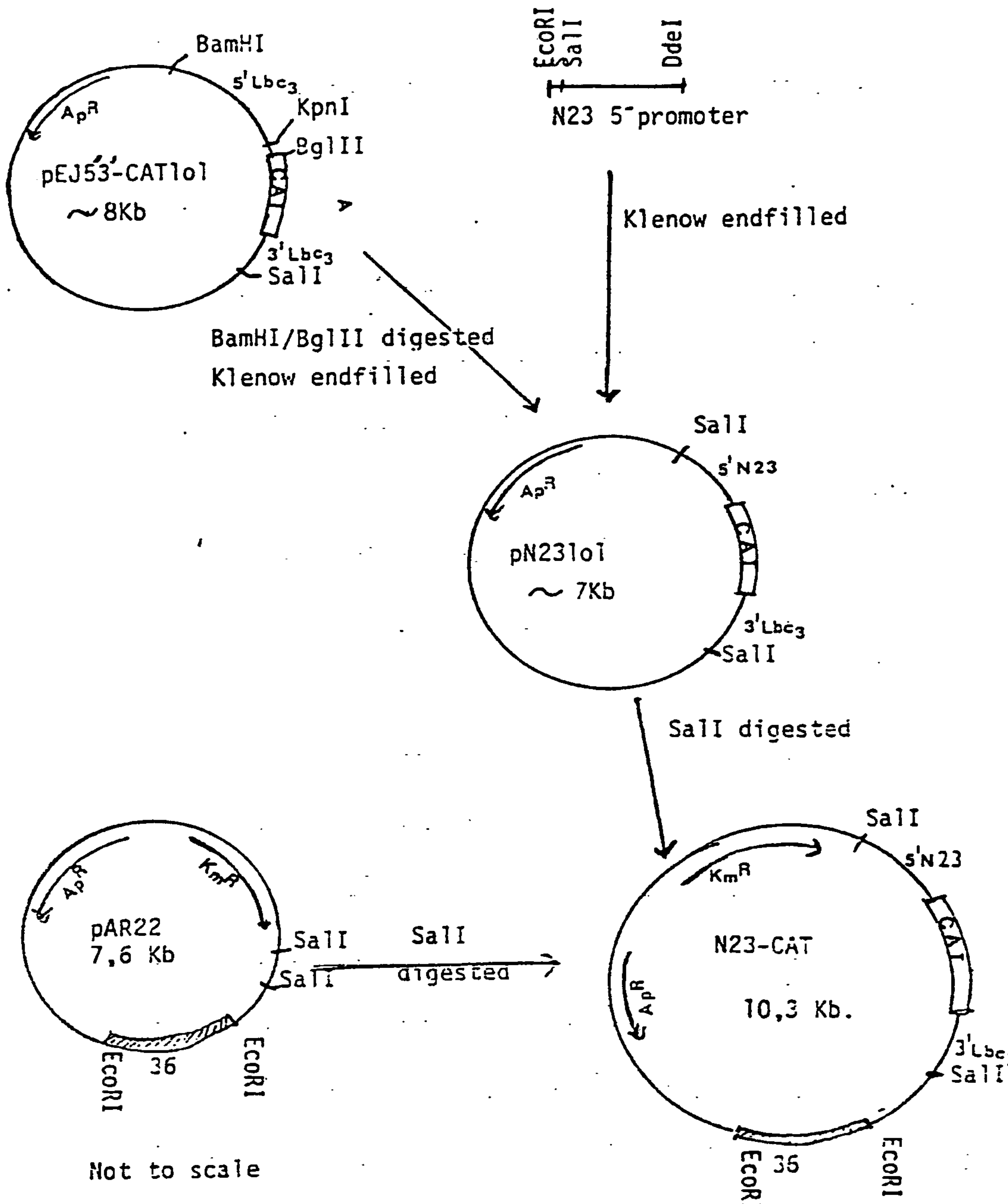
<u>Example 11</u>

<u>Construction of the N23-CAT gene</u>.

The N23 gene was isolated from a soybean DNA library as described in the enclosed paper of Sandal, Bojsen and Marcker. The N23-CAT gene was constructed from the modified Lbc_3-5'-3'-CAT gene carried on plasmid pEJ5'-3'-CAT101 as described in the Applicant's copending application No. 86 11 4704.9 concerning "Expression of Genes in Yeast", and a 1 Kb. EcoRI, DdeI fragment containing the N23 5' promoter region. The position of the EcoRI and DdeI sites in the N23 promoter region is indicated on the DNA sequence shown below. The cloning procedure used is outlined

below. The disclosure of the papers of Sandal et al., the EP application, and the paper of Jensen et al., Nature 321 (12 June 1986), 669-674, including the references cited should be considered incorporated into the present description as a means to amend, illustrate, and clarify it.

The N23-CAT gene was transferred to plants by the same method as the Lbc_3-5'-3'-CAT gene.


BamHI
5'Lbc3
KpnI
BglII
ApR
pEJ53'-CAT1o1
~8Kb
CAT
3'Lbc3
SalI
EcoRI
SalI
DdeI
N23 5' promoter
Klenow endfilled
BamHI/BglII digested
Klenow endfilled
SalI
5'N23
ApR
pN231o1
~ 7Kb
CAT
3'Lbc3
SalI
SalI digested
ApR
KmR
pAR22
7,6 Kb
SalI
SalI
SalI
digested
36
EcoRI
EcoRI
SalI
5'N23
KmR
N23-CAT
CAT
ApR
10,3 Kb.
3'Lbc3
SalI
36
EcoRI
EcoRI

Not to scale

DNA sequence of the 5'-promotor region from the N23 gene

```
         10        20        30        40        50        60        70
GAATTCGAGCTCGCCCGGGGATCGATCCTCTAGAGTCGACCTGCAGCCCAAGCTTGGATCAATCAATTAA
EcoRI                              SalI

         80        90       100       110       120       130       140
TTCTATTGAGACACGATTTGAACAATTTTTACATTATGAGACTATTTTTGGTTTTTTATTTGATCCAAAA

        150       160       170       180       190       200       210
AAATTTAAAGCTTTAGATGATGATGAATTGAANNAATATTGTATTAATNNTGAAAAGTTNNNNNGGTTTA

        220       230       240       250       260       270       280
ATGAATGCTATGATATTGATGGTCTTGATNTATTNNCAGAATTGAAAGTATTAAGAGAAGTGTTAAGAAA

        290       300       310       320       330       340       350
AGAAGTTAGCACACCAATAGAAGTATTGAGTTATATTAAAACTTTAGATTCTTTTCAAATGTTTACATTG

        360       370       380       390       400       410       420
CATATAGAATTTTATTGACAATCCTTATAACAGTTGCTACTGTTGAAAGACGTTCTTCAAAATTAAAATT

        430       440       450       460       470       480       490
ACTTAAATCATATCTAAAATCAACAATGTTACAAGATAGATTGAATGAGTTAGTTATTTTATCTATTGAA

        500       510       520       530       540       550       560
AGTAAAGTGTTAGAATTGTTTGATTATAAAACTCTGATAAATGATTTTGCAGTTAAAAAAACTAGAAGAT

        570       580       590       600       610       620       630
TAATATAAAAATTGATATTTTATATAATATATTAAGTCTCTTTAAAATTCTTGTAAAAAAAGACATTTTT

        640       650       660       670       680       690       700
AAATAATAAAATAAAGCAACTCTTAATTTTAATGAAACATCCCTTTGTTAAACCGAATCTTCCATAATGT

        710       720       730       740       750       760       770
AAAAATTAATGCTTGATGGAAGTTTTTAATTTGTTCTACTCAATACTCAAAGGGTTGTAAATATTTTTTT

        780       790       800       810       820       830       840
TATCATTTATATGTTGTAAATATGAATGCACTAGTAATTAGTTTAATGATAAAATATATTCTACAGATAT

        850       860       870       880       890       900       910
ATTTCTGTCTCTTGGCAACTCGTGAGAATTGAATATATTATAAAGATGAAAGGTCGTTACAATTTTTTTT

        920       930       940       950       960       970       980
AGAATAAATATTTATATACAATTCCTAGATTTTGTTATAAAATTCACATATTGTATGAGTATAAATACAT

        990      1000      1010      1020      1030      1040      1050
GAGCACACACCAAACTAGTCTCAAATTAAGTAAGGTGCTAATTATTAGCGGCTAGCTAAGTAACCAAGTA
                                                       DdeI

ATTAATG
```

Example 12

Organ-specific expression of the soybean N23-CAT gene in root nodules of L.corniculatus and Trifolium repens.

The activity of chloroamphenicol acetyl transferase (CAT) was measured as in example 5 and is given in cpm/μg protein/hrs.

Table a.

	CAT activity	
	N23-CAT transformed L.corniculatus	Untransformed L.corniculatus
Root nodule	86150	0
Root	0	0

Table b.

	CAT activity	
	N23-CAT transformed T.repens	Untransformed T.repens
Root nodule	148000	0
Root	0	0

Table (a) and b) shows the organ-specific expression of the N23-CAT gene in root nodules of L.corniculatus and T.repens. L.corniculatus was inoculated with Rhizobium loti, while T.repens was inoculated with Rhizobium trifolii.

In connection with the invention it has thus been proved that root nodule-specific genes can be expressed organ-specifically upon transfer to other plants, here Lotus corniculatus and Trifolium re-

pens. It has furthermore been proved that the 5' flanking regions comprising the promoter are controlled by the organ-specific regulatory mechanism as the organ-specific control of the Lbc_3-5'-3'-CAT gene in Lotus corniculatus took place at the transcription level. The Lbc_3-5'-3'-CAT gene transferred was thus only transcribed in root nodules of transformed plants and not in other organs such as roots, stems, and leaves.

The expression of the Lbc_3-5'-3'-CAT gene in root nodules of transformed plants also followed the developmental timing known from soybean root nodules. No CAT activity could be detected in roots or small white root nodules (Example 8). A low activity was present in the further developed white distinct nodules, whereas a high activity could be measured in the small pink nodules and mature nodules developed later on.

The organ-specific expression and the correct developmental expression of transferred root nodule-specific genes, here exemplified by the Lbc_3-5'-3'-CAT gene, allows as a particular use a functional expression of root nodule-specific genes also in other plants beyond leguminous plants. When all the root nodule-specific plant genes necessary for the formation of root nodules are transferred from a leguminous plant to a non-root-nodule-forming plant species, the correct organ-specific expression proved above allows production of functionally active, nitrogen-fixing root nodules on this plant upon infecti n by Rhizobium. In this manner these plants can grow without the supply of external

inorganic or organic nitrogen compounds. Root nodule-specific promoters, here exemplified by the Lbc_3 and N23 promoters, must be used in the present case for regulating the expression of the transferred genes.

According to the present invention a root nodule-specific promoter is used for expressing genes. The gene product or function of the gene product improves the function of the root nodule, e.g. by altering the oxygen transport, the metabolism, the nitrogen fixation or the nitrogen absorption.

Root nodules are thus used for the synthesis of biological products improving the plant per se or which can be extracted from the plant later on. A root nodule-specific promoter can be used for expressing a gene. The gene product or compound formed by said gene product constitute the desired product(s).

In connection with the present invention it has furthermore been proved that the soybean Lbc_3 leghemoglobin protein per se, i.e. the Lbc_3 gene product, is present in a high concentration in root nodules of bird's-foot trefoil plants expressing the Lbc_3 code sequence under the control of the Lbc_3 promoter. The latter has been proved by cloning the genomic Lbc_3 gene of the soybean into the integration vector pAR1, said genomic Lbc_3 gene containing the coding sequence, the intervening sequences, and the 5' and 3' flanking sequences. A 3.6 Kb BamHI fragment Lbc_3HH, cf. Example 2, was cloned into the pAR1 plasmid and transferred to

bird's-foot trefoil as stated previously.

The high level of Lbc_3 protein, cf. Example 9, found in transformed root nodules of bird's-foot trefoil and corresponding to the level in soybean root nodules proves an efficient transcription of the Lbc_3 promoter and an efficient processing and translation of Lbc_3mRNA in bird's-foot trefoil.

The high level of the CAT activity present in transformed root nodules is also a result of an efficient translation of mRNA formed from the chimeric Lbc_3 gene. The leader sequence on the Lbc_3 gene is decisive for the translation initiation and must determine the final translation efficiency. This efficiency is of importance for an efficient synthesis of gene products in plants or plant cells. An Lbc_3 or another leghemoglobin leader sequence can thus be used for increasing the final expression level of a predetermined plant promoter. The construction of a DNA fragment comprising a Lb leader sequence as first sequence and an arbitrary promoter as second sequence is a particular use of the invention when the construction is transferred and expressed in plants.

During nodule development around 30 different plant encoded polypeptides (nodulins) are specifically synthesized. Apart from the leghemoglobins, nodulins include nodule-specific forms of uricase (Bergmann et al (1983) EMBO. J. 2, 2333-2339), glutamine synthetase (Cullimore et al (1984) J.Mol. Appl. Genetics 2, 589-599) and sucrose synthase (Morell and C peland (1985) Plant. Physiol. 78,

149-154). The function of most nodulins are, however, at present unknown.

Many nodulin genes have nevertheless been isolated and characterised during the last five years. These include nodulins from several different legumes. Examples of such isolations and characterisations are widespread in the literature such as (Fuller et al (1983) Proc. Natl. Acad.Sci. 80, 2594-2598), (Sengupta-Gopalan et al (1986) Molec. Gen. Genet. 203, 410-420), (Bisseling et al (1985) in Proceedings of the 6th Int. symp. on Nitrogen Fixation, Martinus Nijhoff Publishers pp 53-59.), and (Gebhardt et al (1986) EMBO.J.5, 1429-1435). All of these genes contain nodule-specific regulatory sequences. Such sequences and in fact entire 5' flanking regions and 3' flanking regions can furthermore be synthesized by automated oligonucleotide synthesis knowing the DNA sequences for the Lbc_3 and N23 genes given in this description. Entire nodule-specific genes can also be isolated with known recombinant techniques as described in the above papers and by (Maniatis et al (1982) Molecular cloning. A Laboratory Manual, Cold Spring Harbour Laboratory, New York).

The described method to obtain nodule-specific expression of genes can thus be reconstructed and performed according to the invention by any one skilled in the art of molecular genetics.

The method to obtain nodule-specific expression is not dependent on the <u>A. rhizogenes</u> plant transformation described. Any other plant transformation

system e.g. A. tumefaciens systems, direct gene transfer or microinjection can equally be applied.

The A. rhizogenes system has been used and characterised by a number of scientific groups and is thus well-known from the literature. The characteristics of the system is described in:

Willmitzer et al. (1982), Molec.Gen. Genet. 186, 16-22,

Chilton et al. (1982), Nature 295, 432-434,

Simpson et al. (1986), Plant.Molec.Biol. 6, 493-415,

Tepfer D. (1983), Molecular Genetics of the Bacteria - Plant interaction,

Springer Verlag, Berlin Heidelberg pp 248-258,

White and Nester (1980), J.Bact. 144, 710-720,

Jaynes and Strobel (1981), Int.Rev. of Cytol. Sup. 13, 105-125,

White and Nester (1980), J. Bact. 141, 1134-1141,

Pomponi et al. (1983), Plasmid 10, 119-129, and

Slightom et al. (1986), J. Biol. Chem. 261, 108-121.

The latter two publications describe the restriction map and nucleotide sequence of the *A. rhizogenes* T_L-DNA segment used in the transformation system described here. With this information it is possible to anybody skilled in molecular genetics to use and reconstruct the "intermediate vectors" and the *A. rhizogenes* strains described here.

Claims:

1. A method of expressing genes in plants, parts of plants, and plant cell cultures by introducing into a cell thereof a recombinant DNA segment containing both the gene to be expressed and a 5' flanking region comprising a promoter sequence, and optionally a 3' flanking region, and culturing of the transformed cells in a growth medium, c h a r a c t e r i s e d by using as the recombinant DNA segment a DNA fragment comprising an inducible plant promoter (as defined) from root nodule-specific genes.

2. A method as claimed in claim 1, c h a r a c t e r i s e d by using a DNA fragment comprising an inducible plant promoter (as defined) and being identical with, derived from or comprising 5' flanking regions of root nodule-specific genes.

3. A method as claimed in claim 2, c h a r a c t e r i s e d by using a DNA fragment comprising an inducible plant promoter (as defined) and being identical with, derived from or comprising 5' flanking regions of root nodule-specific genes, said DNA fragment causing an expression of a gene which is induced in root nodules at specific stages of development and as a step of the symbiosis, whereby nitrogen fixation occurs.

4. A method as claimed in claims 1-3 for the expression of root nodule-specific genes, c h a r a c t e r i s e d by using a DNA fragment c mprising an inducible plant promoter (as defined)

from root nodule-specific genes.

5. A method as claimed in claims 1-3 for the expression of genes in leguminous plants, parts of leguminous plants, and leguminous plant cell cultures, c h a r a c t e r i s e d by using a DNA fragment comprising an inducible plant promoter (as defined) from root nodule-specific genes.

6. A method as claimed in claims 1-5, c h a r a c t e r i s e d by the DNA fragment comprising the inducible plant promoter and being identical with, derived from or comprising 5′ flanking regions of leghemoglobin genes.

7. A method as claimed in claim 6, c h a r a c t e r i s e d by the DNA fragment comprising the inducible plant promoter and being identical with, derived from or comprising 5′ flanking regions of soybean leghemoglobin genes.

8. A method as claimed in claim 7, c h a r a c t e r i s e d by the DNA fragment comprising the inducible plant promoter and being identical with, derived from or comprising 5′ flanking regions of the Lba gene with the sequence

```
GAGATACATT ATAATAATCT CTCTAGTGTC TATTTATTAT TTTATCTGGT
GATATATACC TTCTCGTATA CTGTTATTTT TTCAATCTTG TAGATTTACT
TCTTTTATTT TTATAAAAAA GACTTTATTT TTTTAAAAAA AATAAAGTGA
ATTTTGAAAA CATGCTCTTT GACAATTTTC TGTTTCCTTT TTCATCATTG
GGTTAAATCT CATAGTGCCT CTATTCAATA ATTTGGGCTC AATTTAATTA
GTAGAGTCTA CATAAAATTT ACCTTAATAG TAGAGAATAG AGAGTCTTGG
AAAGTTGGTT TTTCTCGAGG AAGAAAGGAA ATGTTAAAAA CTGTGATATT
TTTTTTTTGG ATTAATAGTT ATGTTTATAT GAAAACTGAA AATAAATAAA
CTAACCATAT TAAATTTAGA ACAACACTTC AATTATTTTT TTAATTTGAT
TAATTAAAAA ATTATTTGAT TAAATTTTTT AAAAGATCGT TGTTTCTTCT
TCATCATGCT GATTGACACC CTCCACAAGC CAAGAGAAAC ACATAAGCTT
TGGTTTTCTC ACTCTCCAAG CCCTCTATAT AAACAAATAT TGGAGTGAAG
```

TTGTTGCATA ACTTGCATCG AACAATTAAT AGAAATAACA GAAAATTAAA
AAAGAAATAT G.

9. A method as claimed in claim 7, c h a r a c t e r i s e d by the DNA fragment comprising the inducible plant promoter and being identical with, derived from or comprising 5' flanking regions of the Lbc_1 gene with the sequence:

TTCTCTTAAT ACAATGGAGT TTTTGTTGAA CATACATACA TTTAAAAAAA
AATCTCTAGT GTCTATTTAC CCGGTGAGAA GCCTTCTCGT GTTTTACACA
CTTTAATATT ATTATATCCT CAACCCCACA AAAAAGAATA CTGTTATATC
TTTCCAAACC TGTAGATTTA TTTATTTATT TATTTATTTT TACAAAGGAG
ACTTCAGAAA AGTAATTACA TAAAGATAGT GAACATCATT TTATTTATTA
TAATAAACTT TAAAATCAAA CTTTTTTATA TTTTTTGTTA CCCTTTTCAT
TATTGGGTGA AATCTCATAG TGAAGCCATT AAATAATTTG GGCTCAAGTT
TTATTAGTAA AGTCTGCATG AAATTTAACT TAACAATAGA GAGAGTTTTC
GAAAGGGAGC GAATGTTAAA AAGTGTGATA TTATATTTTA TTTCGATTAA
TAATTATGTT TACATGAAAA CATACAAAAA AATACTTTTA AATTCAGAAT
AATACTTAAA ATATTTATTT GCTTAATTGA TTAACTGAAA ATTATTTGAT
TAGGATTTTG AAAAGATCAT TGGCTCTTCG TCATGCCGAT TGACACCCTC
CACAAGCCAA GAGAAACTTA AGTTGTAAAC TTTCTCACTC CAAGCCTTCT
ATATAAACAT GTATTGGATG TGAAGTTATT GCATAACTTG CATTGAACAA
TAGAAAATAA CAAAAAAAAG TAAAAAAGTA GAAAAGAAAT ATG,

10. A method as claimed in claim 7, c h a r a c t e r i s e d by the DNA fragment comprising the inducible plant promoter and being identical with, derived from or comprising 5' flanking regions of the Lbc_2 gene with the sequence:

TCGAGTTTTT ACTGAACATA CATTTATTAA AAAAAACTCT CTAGTGTCCA
TTTATTCGGC GAGAAGCCTT CTCGTGCTTT ACACACTTTA ATATTATTAT
ATCCCCACCC CCACCAAAAA AAAAAAAACT GTTATATCTT TCCAGTACAT
TTATTTCTTA TTTTTACAAA GGAAACTTCA CGAAAGTAAT TACAAAAAAG
ATAGTGAACA TCATTTTTTT AGTTAAGATG AATTTTAAAA TCACACTTTT
TTATATTTTT TTGTTACCCT TTTCATTATT GGGTGAAATC TCATAGTGAA
ACTATTAAAT AGTTTGGGCT CAAGTTTTAT TAGTAAAGTC TGCATGAAAT
TTAACTTAAT AATAGAGAGA GTTTTGGAAA GGTAACGAAT GTTAGAAAGT
GTGATATTAT TATAGTTTTA TTTAGATTAA TAATTATGTT TACATGAAAA
TTGACAATTT ATTTTTAAAA TTCAGAGTAA TACTTAAATT ACTTATTTAC
TTTAAGATTT TGAAAAGATC ATTTGGCTCT TCATCATGCC GATTGACACC
CTCCACAAGC CAAGAGAAAC TTAAGTTGTA ATTTTTCTAA CTCCAAGCCT
TCTATATAAA CACGTATTGG ATGTGAAGTT GTTGCATAAC TTGCATTGAA
CAATAGAAAT AACAACAAAG AAAATAAGTG AAAAAAGAAA TATG,

11. A method as claimed in claim 7, c h a r a c t e r i s e d by the DNA fragment comprising the inducible plant promoter and being identical with, derived from or comprising 5′ flanking regions of the Lbc_3 gene with the sequence:

```
TATGAAGATT AAAAAATACA CTCATATATA TGCCATAAGA ACCAACAAAA
GTACTATTTA AGAAAAGAAA AAAAAAACCT GCTACATAAT TTCCAATCTT
GTAGATTTAT TTCTTTTATT TTTATAAAGG AGAGTTAAAA AAATTACAAA
ATAAAAATAG TGAACATCGT CTAAGCATTT TTATATAAGA TGAATTTTAA
AAATATAATT TTTTTGTCTA AATCGTATGT ATCTTGTCTT AGAGCCATTT
TTGTTTAAAT TGGATAAGAT CACACTATAA AGTTCTTCCT CCGAGTTTGA
TATAAAAAAA ATTGTTTCCC TTTTGATTAT TGGATAAAAT CTCGTAGTGA
CATTATATTA AAAAAATTAG GGCTCAATTT TTATTAGTAT AGTTTGCATA
AATTTTAACT TAAAAATAGA GAAAATCTGG AAAAGGGACT GTTAAAAAGT
GTGATATTAG AAATTTGTCG GATATATTAA TATTTTATTT TATATGGAAA
CTAAAAAAAT ATATATTAAA ATTTTAAATT CAGAATAATA CTTAAATTAT
TTATTTACTG AAAATGAGTT GATTTAAGTT TTTGAAAAGA TGATTGTCTC
TTCACCATAC CAATTGATCA CCCTCCTCCA ACAAGCCAAG AGAGACATAA
GTTTTATTAG TTATTCTGAT CACTCTTCAA GCCTTCTATA TAAATAAGTA
TTGGATGTGA AGTTGTTGCA TAACTTGCAT TGAACAATTA ATAGAAATAA
CAGAAAAGTA GAAAAGAAAT ATG.
```

12. A method as claimed in claim 7, c h a r a c t e r i s e d by the DNA fragment comprising the inducible plant promoter and being identical with, derived from or comprising 5′ flanking regions of the Lbc_3-5′-3′-CAT gene with the sequence:

```
TATGAAGATT AAAAAATACA CTCATATATA TGCCATAAGA ACCAACAAAA
GTACTATTTA AGAAAAGAAA AAAAAAACCT GCTACATAAT TTCCAATCTT
GTAGATTTAT TTCTTTTATT TTTATAAAGG AGAGTTAAAA AAATTACAAA
ATAAAAATAG TGAACATCGT CTAAGCATTT TTATATAAGA TGAATTTTAA
AAATATAATT TTTTTGTCTA AATCGTATGT ATCTTGTCTT AGAGCCATTT
TTGTTTAAAT TGGATAAGAT CACACTATAA AGTTCTTCCT CCGAGTTTGA
TATAAAAAAA ATTGTTTCCC TTTTGATTAT TGGATAAAAT CTCGTAGTGA
CATTATATTA AAAAAATTAG GGCTCAATTT TTATTAGTAT AGTTTGCATA
AATTTTAACT TAAAAATAGA GAAAATCTGG AAAAGGGACT GTTAAAAAGT
GTGATATTAG AAATTTGTCG GATATATTAA TATTTTATTT TATATGGAAA
CTAAAAAAAT ATATATTAAA ATTTTAAATT CAGAATAATA CTTAAATTAT
TTATTTACTG AAAATGAGTT GATTTAAGTT TTTGAAAAGA TGATTGTCTC
TTCACCATAC CAATTGATCA CCCTCCTCCA ACAAGCCAAG AGAGACATAA
GTTTTATTAG TTATTCTGAT CACTCTTCAA GCCTTCTATA TAAATAAGTA
TTGGATGTGA AGTTGTTGCA TAACTTGCAT TGAACAATTA ATAGAAATAA
CAGAAAAGTA GAATTCTAAA ATG
```

13. A method as claimed in claim 5, c h a r a c t e r i s e d by the DNA fragment comprising the inducible plant promoter and being identical with, derived from or comprising 5' flanking regions of the N23 gene with the sequence:

```
        10        20        30        40        50        60        70
GAATTCGAGCTCGCCCGGGGATCGATCCTCTAGAGTCGACCTGCAGCCCAAGCTTGGATCAATCAATTAA
EcoRI                              SalI
        80        90       100       110       120       130       140
TTCTATTGAGACACGATTTGAACAATTTTTACATTATGAGACTATTTTTGGTTTTTTATTTGATCCAAAA
       150       160       170       180       190       200       210
AAATTTAAAGCTTTAGATGATGATGAATTGAANNAATATTGTATTAATNNTGAAAAGTTNNNNNGGTTTA
       220       230       240       250       260       270       280
ATGAATGCTATGATATTGATGGTCTTGATNTATTNNCAGAATTGAAAGTATTAAGAGAAGTGTTAAGAAA
       290       300       310       320       330       340       350
AGAAGTTAGCACACCAATAGAAGTATTGAGTTATATTAAAACTTTAGATTCTTTTCAAATGTTTACATTG
       360       370       380       390       400       410       420
CATATAGAATTTTATTGACAATCCTTATAACAGTTGCTACTGTTGAAAGACGTTCTTCAAAATTAAAATT
       430       440       450       460       470       480       490
ACTTAAATCATATCTAAAATCAACAATGTTACAAGATAGATTGAATGAGTTAGTTATTTTATCTATTGAA
       500       510       520       530       540       550       560
AGTAAAGTGTTAGAATTGTTTGATTATAAAACTCTGATAAATGATTTTGCAGTTAAAAAAACTAGAAGAT
       570       580       590       600       610       620       630
TAATATAAAAATTGATATTTTATATAATATATTAAGTCTCTTTAAAATTCTTGTAAAAAAAGACATTTTT
       640       650       660       670       680       690       700
AAATAATAAAATAAAGCAACTCTTAATTTTAATGAAACATCCCTTTGTTAAACCGAATCTTCCATAATGT
       710       720       730       740       750       760       770
AAAAATTAATGCTTGATGGAAGTTTTTAATTTGTTCTACTCAATACTCAAAGGGTTGTAAATATTTTTTT
       780       790       800       810       820       830       840
TATCATTTATATGTTGTAAATATGAATGCACTAGTAATTAGTTTAATGATAAAATATATTCTACAGATAT
       850       860       870       880       890       900       910
ATTTCTGTCTCTTGGCAACTCGTGAGAATTGAATATATTATAAAGATGAAAGGTCGTTACAATTTTTTTT
       920       930       940       950       960       970       980
AGAATAAATATTTATATACAATTCCTAGATTTTGTTATAAAATTCACATATTGTATGAGTATAAATACAT
       990      1000      1010      1020      1030      1040      1050
GAGCACACACCAAACTAGTCTCAAATTAAGTAAGGTGCTAATTATTAGCGGCTAGCTAAGTAACCAAGTA
                                                        DdeI
ATTAATG
```

14. A method as claimed in any of the claims 1-13, c h a r a c t e r i s e d by the 3′ flanking region of the genes to be expressed being a 3′ flanking region of root nodule-specific genes of any origin.

15. A method as claimed in claim 14, c h a r - a c t e r i s e d by the 3′ flanking region being of leghemoglobin genes.

16. A method as claimed in claim 14, c h a r - a c t e r i s e d by the 3′ flanking region being of soybean leghemoglobin genes.

17. A method as claimed in claim 16, c h a r - a c t e r i s e d by the 3′ flanking region being of the Lba, Lbc_1, Lbc_2 or Lbc_3 gene with the following sequences, respectively:

Lba

```
                                   1590                                        1620
                        TAA TTA GTA TCT ATT GCA GTA AAG TGT AAT AAA TAA ATC TTG

                                   1650                                        1680
TTT CAC TAT AAA ACT TGT TAC TAT TAG ACA AGG GCC TGA TAC AAA ATG TTG GTT AAA ATA

                                   1710                                        1740
ATG GAA TTA TAT AGT ATT GGA TAA AAA TCT TAA GGT TAA TAT TCT ATA TTT GCG TAG GTT

                                   1770                                        1800
TAT GCT TGT GAA TCA TTA TCG GTA TTT TTT TTC CTT TCT GAT AAT TAA TCG GTA AAT TA

                                   1830                                        1860
ACA AAT AAG TTC AAA ATG ATT TAT ATG TTT CAA AAT TAT TTT AAC AGC AGG TAA AAT GTT

ATT TGG TAC GAA AGC TAA TTC GTC GA
```

Lbc$_1$

```
                                                    1320
             TAA/TT AGG ATC TAC TGC ATT GCC GTA

                                   1350                           1380
AAG TGT AAT AAA TAA ATC TTG TTT CAA CTA AAA CTT GTT ATT AAA CAA GTT CCC TAT ATA

                                   1410                           1440
AAT GTT GTT TAA AAT AAG TAA ATT TCA TTG TAT TGG ATA AAC ACT TTT AAG TTA TAT ATT

                                   1470                           1500
TCC ATA TAT TTA CGT TTG TGA ATC ATA ATC GAT ACT TTA TAA AAA TAA ATT CCA AAT AAT

TTA TAC GTT TTA AAA ATT ATT TT
```

Lbc$_2$

```
                        TAG/GAT CTA CTA TTG CCG TCA AGT
                                                   1140

GTA ATA AAT AAA TTT TGT TTC ACT AAA ACT TGT TAT TAA ACA AGT CCC CGA TAT ATA AAT
                                   1170                                    1200
GTT GGT TAA AAT AAG TAA ATT ATA CGG TAT TGA TAA ACA ATC TTA AGT TTT ATA TAT AGT
                                   1230                                    1260

TCC ATA TAC TAA AGT TTG TGA ATC ATA ATC GA
                                   1290
```

and Lbc$_3$

```
                   TAG/GAT CTA CAA TTG CCT TAA AGT GTA ATA AAT AAA
                           990                                 1020

TAT TAT TTC ACT AAA ACT TGT TAT TAA ACC AAG TTC TCG ATA TAA ATG TTG GTT AAA CTA
                                   1050                                    1080

AGT AAA TTA TAT GGT ATT GGA TAA ACA ATC TTA AGC TT
                                   1110
```

18. A method as claimed in claim 1 of preparing a polypeptide by introducing into a cell of a plant, a part of a plant or a plant cell culture a recombinant plasmid, characterised by using as the recombinant plasmid a plasmid comprising an inducible plant promoter (as defined) f root nodule-specific genes.

19. A DNA fragment comprising an inducible plant promoter (as defined) to be used when carrying out the method as claimed in claims 1-18, c h a r a c t e r i s e d by being identical with, derived from or comprising a 5′ flanking region of root nodule-specific genes of any origin.

20. A DNA fragment as claimed in claim 19, c h a r a c t e r i s e d by being identical with, derived from or comprising a 5′ flanking region of plant leghemoglobin genes.

21. A DNA fragment as claimed in claim 20, c h a r a c t e r i s e d by being identical with, derived from or comprising a 5′ flanking region of soybean leghemoglobin genes.

22. A DNA fragment as claimed in claim 21, c h a r a c t e r i s e d by being identical with, derived from or comprising a 5′ flanking region of the Lba gene with the sequence:

```
GAGATACATT ATAATAATCT CTCTAGTGTC TATTTATTAT TTTATCTGGT
GATATATACC TTCTCGTATA CTGTTATTTT TTCAATCTTG TAGATTTACT
TCTTTTATTT TTATAAAAAA GACTTTATTT TTTTAAAAAA AATAAAGTGA
ATTTTGAAAA CATGCTCTTT GACAATTTTC TGTTTCCTTT TTCATCATTG
GGTTAAATCT CATAGTGCCT CTATTCAATA ATTTGGGCTC AATTTAATTA
GTAGAGTCTA CATAAAATTT ACCTTAATAG TAGAGAATAG AGAGTCTTGG
AAAGTTGGTT TTTCTCGAGG AAGAAAGGAA ATGTTAAAAA CTGTGATATT
TTTTTTTTGG ATTAATAGTT ATGTTTATAT GAAAACTGAA AATAAATAAA
CTAACCATAT TAAATTTAGA ACAACACTTC AATTATTTTT TTAATTTGAT
TAATTAAAAA ATTATTTGAT TAAATTTTTT AAAAGATCGT TGTTTCTTCT
TCATCATGCT GATTGACACC CTCCACAAGC CAAGAGAAAC ACATAAGCTT
TGGTTTTCTC ACTCTCCAAG CCCTCTATAT AAACAAATAT TGGAGTGAAG
TTGTTGCATA ACTTGCATCG AACAATTAAT AGAAATAACA GAAAATTAAA
AAAGAAATAT G,
```

23. A DNA fragment as claimed in claim 21, c h a r a c t e r i s e d by being identical with, derived from or comprising a 5′ flanking region of the Lbc_1 gene with the sequence:

```
TTCTCTTAAT ACAATGGAGT TTTTGTTGAA CATACATACA TTTAAAAAAA
AATCTCTAGT GTCTATTTAC CCGGTGAGAA GCCTTCTCGT GTTTTACACA
CTTTAATATT ATTATATCCT CAACCCCACA AAAAAGAATA CTGTTATATC
TTTCCAAACC TGTAGATTTA TTTATTTATT TATTTATTTT TACAAAGGAG
ACTTCAGAAA AGTAATTACA TAAAGATAGT GAACATCATT TTATTTATTA
TAATAAACTT TAAAATCAAA CTTTTTTATA TTTTTTGTTA CCCTTTTCAT
TATTGGGTGA AATCTCATAG TGAAGCCATT AAATAATTTG GGCTCAAGTT
TTATTAGTAA AGTCTGCATG AAATTTAACT TAACAATAGA GAGAGTTTTC
GAAAGGGAGC GAATGTTAAA AAGTGTGATA TTATATTTTA TTTCGATTAA
TAATTATGTT TACATGAAAA CATACAAAAA AATACTTTTA AATTCAGAAT
AATACTTAAA ATATTTATTT GCTTAATTGA TTAACTGAAA ATTATTTGAT
TAGGATTTTG AAAAGATCAT TGGCTCTTCG TCATGCCGAT TGACACCCTC
CACAAGCCAA GAGAAACTTA AGTTGTAAAC TTTCTCACTC CAAGCCTTCT
ATATAAACAT GTATTGGATG TGAAGTTATT GCATAACTTG CATTGAACAA
TAGAAAATAA CAAAAAAAAG TAAAAAAGTA GAAAAGAAAT ATG,
```

24. A DNA fragment as claimed in claim 21, c h a r a c t e r i s e d by being identical with, derived from or comprising a 5′ flanking region of the Lbc_2 gene with the sequence:

```
TCGAGTTTTT ACTGAACATA CATTTATTAA AAAAAACTCT CTAGTGTCCA
TTTATTCGGC GAGAAGCCTT CTCGTGCTTT ACACACTTTA ATATTATTAT
ATCCCCACCC CCACCAAAAA AAAAAAAACT GTTATATCTT TCCAGTACAT
TTATTTCTTA TTTTTACAAA GGAAACTTCA CGAAAGTAAT TACAAAAAAG
ATAGTGAACA TCATTTTTTT AGTTAAGATG AATTTTAAAA TCACACTTTT
TTATATTTTT TTGTTACCCT TTTCATTATT GGGTGAAATC TCATAGTGAA
ACTATTAAAT AGTTTGGGCT CAAGTTTTAT TAGTAAAGTC TGCATGAAAT
TTAACTTAAT AATAGAGAGA GTTTTGGAAA GGTAACGAAT GTTAGAAAGT
GTGATATTAT TATAGTTTTA TTTAGATTAA TAATTATGTT TACATGAAAA
TTGACAATTT ATTTTTAAAA TTCAGAGTAA TACTTAAATT ACTTATTTAC
TTTAAGATTT TGAAAAGATC ATTTGGCTCT TCATCATGCC GATTGACACC
CTCCACAAGC CAAGAGAAAC TTAAGTTGTA ATTTTTCTAA CTCCAAGCCT
TCTATATAAA CACGTATTGG ATGTGAAGTT GTTGCATAAC TTGCATTGAA
CAATAGAAAT AACAACAAAG AAAATAAGTG AAAAAAGAAA TATG,
```

25. A DNA fragment as claimed in claim 21, c h a r a c t e r i s e d by being identical with, derived from or comprising a 5′ flanking region of

the Lbc_3 gene with the sequence:

```
TATGAAGATT AAAAAATACA CTCATATATA TGCCATAAGA ACCAACAAAA
GTACTATTTA AGAAAAGAAA AAAAAAACCT GCTACATAAT TTCCAATCTT
GTAGATTTAT TTCTTTTATT TTTATAAAGG AGAGTTAAAA AAATTACAAA
ATAAAAATAG TGAACATCGT CTAAGCATTT TTATATAAGA TGAATTTTAA
AAATATAATT TTTTTGTCTA AATCGTATGT ATCTTGTCTT AGAGCCATTT
TTGTTTAAAT TGGATAAGAT CACACTATAA AGTTCTTCCT CCGAGTTTGA
TATAAAAAAA ATTGTTTCCC TTTTGATTAT TGGATAAAAT CTCGTAGTGA
CATTATATTA AAAAAATTAG GGCTCAATTT TTATTAGTAT AGTTTGCATA
AATTTTAACT TAAAAATAGA GAAAATCTGG AAAAGGGACT GTTAAAAAGT
GTGATATTAG AAATTTGTCG GATATATTAA TATTTTATTT TATATGGAAA
CTAAAAAAAT ATATATTAAA ATTTTAAATT CAGAATAATA CTTAAATTAT
TTATTTACTG AAAATGAGTT GATTTAAGTT TTTGAAAAGA TGATTGTCTC
TTCACCATAC CAATTGATCA CCCTCCTCCA ACAAGCCAAG AGAGACATAA
GTTTTATTAG TTATTCTGAT CACTCTTCAA GCCTTCTATA TAAATAAGTA
TTGGATGTGA AGTTGTTGCA TAACTTGCAT TGAACAATTA ATAGAAATAA
CAGAAAAGTA GAAAAGAAAT ATG.
```

26. A DNA fragment as claimed in claim 21, c h a r a c t e r i s e d by the DNA fragment comprising the inducible plant promoter being identical with, derived from or comprising 5' flanking regions of Lbc_3-5'-3'-CAT gene with the sequence:

```
TATGAAGATT AAAAAATACA CTCATATATA TGCCATAAGA ACCAACAAAA
GTACTATTTA AGAAAAGAAA AAAAAAACCT GCTACATAAT TTCCAATCTT
GTAGATTTAT TTCTTTTATT TTTATAAAGG AGAGTTAAAA AAATTACAAA
ATAAAAATAG TGAACATCGT CTAAGCATTT TTATATAAGA TGAATTTTAA
AAATATAATT TTTTTGTCTA AATCGTATGT ATCTTGTCTT AGAGCCATTT
TTGTTTAAAT TGGATAAGAT CACACTATAA AGTTCTTCCT CCGAGTTTGA
TATAAAAAAA ATTGTTTCCC TTTTGATTAT TGGATAAAAT CTCGTAGTGA
CATTATATTA AAAAAATTAG GGCTCAATTT TTATTAGTAT AGTTTGCATA
AATTTTAACT TAAAAATAGA GAAAATCTGG AAAAGGGACT GTTAAAAAGT
GTGATATTAG AAATTTGTCG GATATATTAA TATTTTATTT TATATGGAAA
CTAAAAAAAT ATATATTAAA ATTTTAAATT CAGAATAATA CTTAAATTAT
TTATTTACTG AAAATGAGTT GATTTAAGTT TTTGAAAAGA TGATTGTCTC
TTCACCATAC CAATTGATCA CCCTCCTCCA ACAAGCCAAG AGAGACATAA
GTTTTATTAG TTATTCTGAT CACTCTTCAA GCCTTCTATA TAAATAAGTA
TTGGATGTGA AGTTGTTGCA TAACTTGCAT TGAACAATTA ATAGAAATAA
CAGAAAAGTA GAATTCTAAA ATG
```

27. A DNA fragment as claimed in claim 19, c h a r a c t e r i s e d by being identical with,

derived from or comprising 5' flanking regions of the N23 gene with the sequence:

```
         10        20        30        40        50        60        70
GAATTCGAGCTCGCCCGGGGATCGATCCTCTAGAGTCGACCTGCAGCCCAAGCTTGGATCAATCAATTAA
EcoRI                              SalI

         80        90       100       110       120       130       140
TTCTATTGAGACACGATTTGAACAATTTTTACATTATGAGACTATTTTTGGTTTTTTATTTGATCCAAAA

        150       160       170       180       190       200       210
AAATTTAAAGCTTTAGATGATGATGAATTGAANNAATATTGTATTAATNNTGAAAAGTTNNNNNGGTTTA

        220       230       240       250       260       270       280
ATGAATGCTATGATATTGATGGTCTTGATNTATTNNCAGAATTGAAAGTATTAAGAGAAGTGTTAAGAAA

        290       300       310       320       330       340       350
AGAAGTTAGCACACCAATAGAAGTATTGAGTTATATTAAAACTTTAGATTCTTTTCAAATGTTTACATTG

        360       370       380       390       400       410       420
CATATAGAATTTTATTGACAATCCTTATAACAGTTGCTACTGTTGAAAGACGTTCTTCAAAATTAAAATT

        430       440       450       460       470       480       490
ACTTAAATCATATCTAAAATCAACAATGTTACAAGATAGATTGAATGAGTTAGTTATTTTATCTATTGAA

        500       510       520       530       540       550       560
AGTAAAGTGTTAGAATTGTTTGATTATAAACTCTGATAAATGATTTTGCAGTTAAAAAAAACTAGAAGAT

        570       580       590       600       610       620       630
TAATATAAAAATTGATATTTTATATAATATATTAAGTCTCTTTAAAATTCTTGTAAAAAAAGACATTTTT

        640       650       660       670       680       690       700
AAATAATAAAATAAAGCAACTCTTAATTTTAATGAAACATCCCTTTGTTAAACCGAATCTTCCATAATGT

        710       720       730       740       750       760       770
AAAAATTAATGCTTGATGGAAGTTTTTAATTTGTTCTACTCAATACTCAAAGGGTTGTAAATATTTTTTT

        780       790       800       810       820       830       840
TATCATTTATATGTTGTAAATATGAATGCACTAGTAATTAGTTTAATGATAAAATATATTCTACAGATAT

        850       860       870       880       890       900       910
ATTTCTGTCTCTTGGCAACTCGTGAGAATTGAATATATTATAAAGATGAAAGGTCGTTACAATTTTTTTT

        920       930       940       950       960       970       980
AGAATAAATATTTATATACAATTCCTAGATTTTGTTATAAAATTCACATATTGTATGAGTATAAATACAT

        990      1000      1010      1020      1030      1040      1050
GAGCACACACCAAACTAGTCTCAAATTAAGTAAGGTGCTAATTATTAGCGGCTAGCTAAGTAACCAAGTA
                                                         DdeI

ATTAATG
```

28. A plasmid which can be used when carrying

out the method as claimed in claims 1-18, c h a r a c t e r i s e d by comprising a DNA fragment as claimed in any of the claims 19-27.

29. A plasmid as claimed in claim 28, c h a r a c t e r i s e d by being pAR29.

30. A plasmid as claimed in claim 28, c h a r a c t e r i s e d by being pAR30.

31. A plasmid as claimed in claim 28, c h a r a c t e r i s e d by being pAR11.

32. A plasmid as claimed in claim 28, c h a r a c t e r i s e d by being N23-CAT.

33. A transformant <u>Agrobacterium rhizogenes</u> 15834-strain which can be used when carrying out the method as claimed in any of the claims 1 to 18, c h a r a c t e r i s e d by the bacterium strain being transformed by a plasmid according to any of the preceding claims 28 to 32.

34. A transformant <u>Agrobacterium rhizogenes</u> 15834-strain which can be used when carrying out the method as claimed in any of the claims 1 to 18, c h a r a c t e r i s e d by the bacterium strain being transformed by pAR29 and being named AR1127.

35. A transformant <u>Agrobacterium rhizogenes</u> 15834-strain which can be used when carrying out the method as claimed in any of the claims 1 to 18, c h a r a c t e r i s e d by the bacterium strain being transformed by pAR30 and being named AR1134.

36. A transformant <u>Agrobacterium rhizogenes</u> 15834-strain which can be used when carrying out the method as claimed in any of the claims 1 to 18, c h a r a c t e r i s e d by the bacterium strain being transformed by pAR11 and being named AR1000.

37. A transformant <u>Agrobacterium rhizogenes</u> 15834-strain which can be used when carrying out the method as claimed in any of the claims 1 to 18, c h a r a c t e r i s e d by the bacterium strain being transformed by N23-CAT and being named AR204-N23-CAT.

38. Plants, parts of plants and plant cells, particularly of the family Leguminosae, obtainable by transformation with a recombinant DNA segment, fragment or plasmid according to any one of the claims 1 to 37.

16a

PCT

世界知的所有権機関
国　際　事　務　局
特許協力条約に基づいて公開された国際出願

(51) 国際特許分類6		(11) 国際公開番号	WO95/18222
C12N 15/00, 5/00, A01H 1/00, 5/00, //C12N 9/00	A1	(43) 国際公開日	1995年7月6日(06.07.95)

(21) 国際出願番号　PCT/JP94/02288
(22) 国際出願日　1994年12月28日(28.12.94)

(30) 優先権データ
特願平5/352858　1993年12月28日(28.12.93)　JP

(71) 出願人（米国を除くすべての指定国について）
麒麟麦酒株式会社(KIRIN BEER KABUSHIKI KAISHA)[JP/JP]
〒104 東京都中央区新川二丁目１０番1号 Tokyo, (JP)
(72) 発明者；および
(75) 発明者／出願人（米国についてのみ）
西沢治(NISHIZAWA, Osamu)[JP/JP]
戸栗敏博(TOGURI, Toshihiro)[JP/JP]
〒236 神奈川県横浜市金沢区福浦1-13-5
麒麟麦酒株式会社 基盤技術研究所内 Kanagawa, (JP)

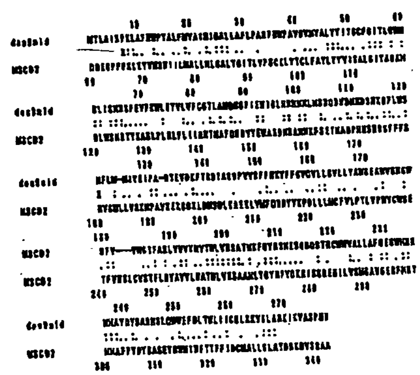
(74) 代理人
弁理士　平木祐輔, 外(HIRAKI, Yusuke et al.)
〒105 東京都港区虎ノ門二丁目7番7号
虎ノ門中田ビル2F Tokyo, (JP)

(81) 指定国
AM, AU, BB, BG, BR, BY, CA, CN, CZ, FI, GE, HU, JP, KG, KR, KZ, LK, LT, LV, MD, MG, MN, NO, NZ, PL, RO, RU, SI, SK, TJ, TT, UA, US, UZ, VN, 欧州特許(AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI特許(BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO特許(KE, MW, SD, SZ).

添付公開書類　国際調査報告書

(54) Title : GENE FOR FATTY ACID DESATURASE, VECTOR CONTAINING SAID GENE, PLANT CONTAINING SAID GENE TRANSFERRED THEREINTO, AND PROCESS FOR CREATING SAID PLANT

(54) 発明の名称　脂肪酸の不飽和化酵素遺伝子、当該遺伝子を含むベクター、当該遺伝子が導入された植物及びその作出方法



(57) Abstract

A gene coding for a protein having the activity of desaturating the Δ9-position of a fatty acid bound to a lipid; a vector containing a polynucleotide containing the whole or part f said gene; a plant cell containing, transferred thereinto, a polynucleotide containing the whole or part of a gene coding for a protein having the activity of desaturating the Δ9-position of a fatty acid bound to a lipid; a process for creating a plant which comprises differentiating said plant cells and regenerating the plant body; and a plant containing, transferred thereinto, a polynucleotide c ntaining the whole or part of a gene coding for protein having the activity of desaturating the Δ9-position of a fatty acid bound to a lipid.

（５７）要約

脂質に結合した脂肪酸のΔ９位を不飽和化する活性を有するタンパク質をコードする遺伝子；当該遺伝子又は当該遺伝子の一部を含むポリヌクレオチドを含むベクター；脂質に結合した脂肪酸のΔ９位を不飽和化する活性を有するタンパク質をコードする遺伝子又は当該遺伝子の一部を含むポリヌクレオチドが導入された植物細胞；前記植物細胞を分化させて植物体を再生させることを特徴とする植物の作出方法；および、脂質に結合した脂肪酸のΔ９位を不飽和化する活性を有するタンパク質をコードする遺伝子又は当該遺伝子の一部を含むポリヌクレオチドが導入された植物。

明 細 書

脂肪酸の不飽和化酵素遺伝子、当該遺伝子を含むベクター、当該遺伝子が導入された植物及びその作出方法

技術分野

本発明は、脂質に結合した脂肪酸のΔ9位を不飽和化する活性を有するタンパク質（以下、Δ9位不飽和化酵素という）をコードする遺伝子、当該遺伝子を含むベクター、当該遺伝子が導入された植物およびその作出方法に関するものである。

背景技術

生物の生体膜を構成する脂質である膜脂質は外界温度の低下に伴って、液晶状態から固体状態へと変化（相分離）する。そして、かかる相分離に伴い生体膜の性質が変化する。すなわち、膜脂質が固体状態では物質透過の選択性がなくなるため、生体膜が本来の機能を果たせなくなり、その結果細胞に傷害（低温傷害）が生ずると考えられている。

液晶状態から固体状態あるいはその逆に変化する温度である膜脂質の相転移温度は、主に脂質に結合している脂肪酸アシル基の不飽和度（炭素鎖中の二重結合の数）によって決定付けられる。すなわち、結合している脂肪酸アシル基が二つとも飽和脂肪酸である場合、この脂質分子種の相転移温度は室温よりも高いが、結合した脂肪酸アシル基に二重結合を少なくとも1個持つような脂質分子種の相転移温度は、ほぼ0℃以下である（Santaren, J.F.et al., Biochim. Biophys. Acta, 687:231, 1982）。

なお、一般に脂肪酸の二重結合の位置は、そのカルボキシル基末端から二重結合のある炭素までの炭素数をΔ（デルタ）に続いて示す。また、二重結合の総数は全炭素数の後にコロンに続いて記載する。例えば、リノール酸は18:2Δ9,12と記述され、その構造は

$$CH_3(CH_2)_4\ CH=CHCH_2\ CH=CH\ (CH_2)_7\ COOH$$

である。また、二重結合の位置をω（オメガ）に続いて記載する場合があるが、

これは脂肪酸のメチル基末端から二重結合のある炭素までの炭素数を示している。

高等植物の膜脂質の中で、飽和分子種が比較的多いのはホスファチジルグリセロール（ＰＧ）のみであり、植物の低温傷害の起因がＰＧの相転移によること（Murata, N. et al., Plant Cell Physiol., 23:1071, 1982; Roughan, P. G., Plant Physiol., 77:740, 1985 ）、またＰＧの分子種組成が葉緑体に存在するグリセロール－３－リン酸アシルトランスフェラーゼ（以下ATase）の基質選択性によって決められていること（Frentzen, M. et al., Eur. J. Biochem., 129:629, 1983; Murata, N., Plant Cell Physiol., 24:81, 1983; Frentzen, M. et al., Plant Cell Physiol., 28:1195, 1988）が強く示唆されていた。

これらの仮定に基づき西澤らは、低温に強い植物のシロイヌナズナから取得したATase遺伝子をタバコに導入・発現することによりＰＧの飽和分子種含量を下げ、タバコを野生株よりも低温に対して強くすることができることを示した（ＰＣＴ特許出願：PCT/JP92/00024, 1992）。しかし、ATaseは元の植物中にも存在し、かりに外来のATaseを植物中で大量発現させたとしても、内在性のATaseと競合しあうことは避けられず、外来のATaseの効果が希釈される可能性は否めない。例えば、作成した形質転換タバコのうちシロイヌナズナのATaseを最も大量に発現しているクローンの葉のＰＧの飽和分子種含量は約28％でありタバコ野生株よりも約８％少ないが、シロイヌナズナ野生株よりも約８％多かった（ＰＣＴ特許出願：PCT/JP92/00024, 1992）。

さらに、一般にプラスチドで作られるアシル-ACP は主に16:0-ACPと18:1-ACPであり、またそれらの割合はほぼ等量であると考えられているが、組織によっては16:0-ACPや18:0-ACPの割合が18:1-ACPより高いことも考えられる（Toriyama, S. et al., Plant Cell Physiol., 29:615, 1988）。このような組織では外来のATaseによって飽和分子種含量を充分に減少させることが困難であるとも考えられる。

ところで、光合成細菌のシアノバクテリア（ラン藻）の膜脂質の組成は、高等植物の葉緑体を構成している膜系の脂質組成と類似している（Murata, N. et al., in“The Biochemistry of Plants”, Academic Press, 1987）。またラン藻では、膜脂質に結合した脂肪酸の不飽和度は、脂質に結合した脂肪酸を不飽和化する酵素によって制御されている。そして、脂質に結合した脂肪酸に二重結合を１つし

か入れられないAnacystis nidulans（別名 Synechococcus PCC 7942)は低温感受性であるが（Ono,T.et al.,Plant Physiol.,67:176,1981)、２つ以上入れられるSynechocystis PCC6803 は低温耐性であることが知られていた（Wada,H.et al.,Plant Cell Physiol.,30:971,1989）。

また、ラン藻における脂肪酸の不飽和化酵素は、すべて脂質を基質とし、脂質に結合した脂肪酸に二重結合を導入する。従って、ラン藻は16:0/16:0-および18: 0/16:0- の飽和分子種からなる膜脂質のＰＧ、ＳＱＤＧ、ＭＧＤＧおよびＤＧＤＧの脂肪酸にcis-型の二重結合を導入することが可能である（Murata,N.et al.,in "The Biochemistry of Plants",Academic Press, 1987）。かかる点は脂肪酸不飽和化酵素として、ステアロイル-ACP（18:0-ACP）のΔ９位に二重結合を導入する酵素を有し、一旦16:0/16:0-（および、わずかに存在する18:0/16:0-）の飽和分子種からなるＰＧおよびＳＱＤＧが合成されると、それらの脂肪酸に決してcis-型の二重結合を導入しない高等植物と大きく異なる点である。

現在、Synechocystis PCC6803 のΔ12位不飽和化酵素遺伝子をAnacystis nidulansに導入・発現させることにより本来Anacystis nidulansには存在しない16:2Δ9,12および18:2Δ9,12を生産させることが可能であり、結果として本来低温感受性であるAnacystis nidulansを低温耐性へと転換可能であることが示されている（Wada,H.et al.,Nature,347:200,1990）。

なお、これまでにラン藻の不飽和化酵素のうちΔ６位（Reddy,A.S.et al.,Plant Mol.Biol.,27:293,1993）およびΔ12位（Wada,H.et al.,Nature,347:200,1990)不飽和化酵素の遺伝子が取得されている。しかし、Δ９位に二重結合が導入されていなければ、Δ６位およびΔ12位不飽和化酵素は、それぞれΔ６位とΔ12位を不飽和化することはできない。また、Δ９位とΔ12位がともに不飽和化されていなければΔ15位不飽和化酵素はΔ15位を不飽和化することはできない。従って、脂肪酸のΔ９位を不飽和化する酵素の遺伝子を高等植物に導入し発現させれば、高等植物における飽和分子種含量を低下させ、その結果として該高等植物を低温耐性にすることができるはずである。しかしながら、現在まで、脂肪酸のΔ９位を不飽和化する酵素の遺伝子は得られていなかった。

従って、本発明は、脂肪酸のΔ９位を不飽和化する酵素の遺伝子およびその一

部を含むポリヌクレオチドを提供することを目的とする。

また、本発明は、脂肪酸のΔ９位を不飽和化する酵素の遺伝子またはその一部を含むポリヌクレオチドを含むベクターを提供することも目的とする。

さらに、本発明は、脂肪酸のΔ９位を不飽和化する酵素の遺伝子またはその一部を含むポリヌクレオチドが導入された植物細胞および植物を提供することも目的とする。

発明の開示

上記目的を達成するため、本発明者は、Anacystis 属に属するラン藻のゲノムＤＮＡからΔ９位不飽和化酵素をコードする遺伝子をクローニングし、該遺伝子を組み込んだベクターＤＮＡを得た後、該ベクターＤＮＡで植物細胞を形質転換し、これを分化させて植物体を再生させることにより、植物に低温耐性を付与することに成功し、本発明を完成させるに至った。すなわち、本発明は、以下の事項を要旨とするものである。

(1) 脂質に結合した脂肪酸のΔ９位を不飽和化する活性を有するタンパク質をコードする遺伝子。

(2) 脂質に結合した脂肪酸のΔ９位を不飽和化する活性を有するタンパク質が実質的に配列番号４に記載されたアミノ酸配列を有するものである(1)に記載の遺伝子又は当該遺伝子の一部を含むポリヌクレオチド。

(3) 脂質に結合した脂肪酸のΔ９位を不飽和化する活性を有するタンパク質をコードする遺伝子が配列番号３に記載の塩基配列を含むＤＮＡ鎖である(1)に記載の遺伝子又は当該遺伝子の一部を含むポリヌクレオチド。

(4) (1)乃至(3)のいずれかに記載の遺伝子又は当該遺伝子の一部を含むポリヌクレオチドを含むベクター。

(5) (1)乃至(3)のいずれかに記載の遺伝子又は当該遺伝子の一部を含むポリヌクレオチドが導入された植物細胞。

(6) (5)に記載の植物細胞を分化させて植物体を再生させることを特徴とする植物の作出方法。

(7) (1)乃至(3)のいずれかに記載の遺伝子又は当該遺伝子の一部を含むポリヌクレオチドレ

オチドが導入された植物。

図面の簡単な説明

第１図は、des 9 var 断片がコードするアミノ酸配列とマウスのステアロイル－ＣｏＡ不飽和化酵素（ＭＳＣＤ２）のアミノ酸配列の比較を示す。図中で両者が同一のアミノ酸の場合は：、性質が類似したアミノ酸の場合は・を付け比較した。Ｘは、その間での相同性が高い範囲を示す。

第２図は、des 9 var 断片をプローブとして、Anacystis nidulans のゲノムＤＮＡをサザン分析したオートラジオグラムを示す電気泳動写真である。

第３図は、λ５、λ15およびｐ15Ｘのインサート DＮＡ断片の相互関係を示す。太い矢印はタンパク質をコードしている部分と方向を、細い矢印はシーケンスを決定した部位とその方向を示す。

第４図は、des 9 nidとマウスのステアロイル－ＣｏＡ不飽和化酵素（ＭＳＣＤ２）のアミノ酸配列の比較を示す。アミノ酸配列の比較は第１図と同様にして行った。

第５図は、植物体レベルでの形質転換タバコに対する低温処理の影響を示す生物の形態の写真である。左は不飽和化酵素遺伝子を導入したタバコを低温処理した結果を、右は対照としてpBI121を導入したタバコを低温処理した結果を示す。

発明の実施するための最良の形態

本発明にいう、Δ９位不飽和化酵素は、上記「従来の技術」等に記載したごとく本来ラン藻に存在する酵素である。Δ９位不飽和化酵素の化学構造は、マウス（Kaestner, K. H. et al., J. Biol. Chem., 264:14755, 1989)、ラット（Mihara, K., J. Biochem., 108:1022, 1990）及び酵母（Stukey, J. E. et al., J. Biol. Chem., 265:20144, 1990 ）のステアロイル－ＣｏＡ不飽和化酵素の化学構造と局所的に類似しているが全体的には大きく異なる。また既知のラン藻の脂質に結合した脂肪酸のΔ６位およびΔ12位の不飽和化酵素及び高等植物の脂質に結合した脂肪酸のω３位の不飽和化酵素（Yadav, N. S. et al., Plant Physiol., 103:467, 1993）の化学構造とは全く類似していない。本発明遺伝子を

天然素材から製造する場合は、ラン藻を原材料として使用するとよい。ここで用いられるラン藻は特に限定されず、例えばAnacystis属、Synechocystis属、Anabaena属等に属するラン藻を挙げることができる。なお、以下の理由により、高等植物の飽和分子種を不飽和化するためにはAnacystis型（Murata,N. et al., Plant Cell Physiol., 33: 933, 1992 。この文献で言うグループ１型のラン藻）のΔ９位不飽和化酵素の方がAnabaena型及びSynechocystis型の酵素よりも好ましい。

すなわち、Synechocystis PCC6803 とAnabaena variabilisでは、その膜脂質のほとんどがsn-1とsn-2にそれぞれ炭素数18の脂肪酸（Ｃ18）と炭素数16の脂肪酸（Ｃ16）を結合している（Sato,N. et al., Biochim. Biophys. Acta, 710: 279, 1982; Wada,H. et al., Plant Cell Physiol., 30:971, 1989 ）のに対して、Anacystis nidulansではほとんどがsn-1とsn-2ともにＣ16を結合している（Bishop, D. G. et al., Plant Cell Physiol., 27:1593, 1986）。従って、AnabaenaとSynechocystisのΔ９位不飽和化酵素は主に18:0/16:0-の分子種を基質としてsn-1の18:0を18:1Δ９に不飽和化する活性を有すると思われる。これに対してAnacystisのΔ９位不飽和化酵素は主に16:0/16:0-の分子種を基質としてsn-1の16:0を16:1Δ９に不飽和化する活性を有すると思われる。さらに、高等植物に多く見られる飽和分子種が16:0/16:0-であることから、高等植物の飽和分子種を不飽和化するためにはAnacystis型のΔ９位不飽和化酵素のほうがAnabaenaおよびSynechocystis型の酵素より適切である。

本発明遺伝子は後述する実施例に示すように、実質的に配列番号４に記載されたアミノ酸配列を有するΔ９位不飽和化酵素をコードするものを含み、縮重コドンにおいてのみ異なっていて同一のポリペプチドをコードすることのできる縮重異性体を含むものである。本発明遺伝子は、主にＤＮＡ鎖としての具体的形態を有する。なお、「実質的に配列番号４に記載されたアミノ酸配列」とは、配列番号４に記載されたアミノ酸配列に加えて、Δ９位不飽和化酵素活性を有するかぎり、配列番号４に記載されたアミノ酸配列の一部に欠失、置換、付加などがあってもよいアミノ酸配列を含むものである。

本発明遺伝子は、上記ラン藻細胞から通常公知の手法を用いて製造することが

できる。

すなわち、ラン藻細胞を培養して集積し、当該ラン藻細胞からエタノール沈澱法等の通常公知の手法によりゲノムＤＮＡを調製し、当該ゲノムＤＮＡを基にした遺伝子ライブラリーを調製し、当該ライブラリーより所望の遺伝子を含むクローンを選抜し、これを増幅することで製造することが可能である。

ここで用いる遺伝子ライブラリー作成用ベクターとしては、当該ベクターとして通常用いられるものを挙げることができる。具体的には、λDASH II (Stratagene)等のファージ；pWE15 (Stratagene)等のコスミド；pBluescript II(Stratagene)等のファージミド等を挙げることができる。上記ベクターへの具体的な遺伝子導入方法は、それぞれのベクターに応じた通常公知の方法を用いることができる。

このようにして調製した遺伝子ライブラリーから本発明遺伝子が導入されたクローンを選抜する。

当該選抜方法としては通常公知の選抜方法、例えば抗体によるプラークハイブリダイゼーション法若しくはコロニーハイブリダイゼーション法等の免疫学的方法又はヌクレオチドプローブによるプラークハイブリダイゼーション法若しくはコロニーハイブリダイゼーション法等を用いることができる。なお、上記ヌクレオチドプローブの選択基準として、本発明遺伝子に類似すると推測される塩基配列の一部（例えば、第１図のMSCD2のアミノ酸配列番号260から295の一部を塩基配列に読みかえたもの）をプローブとして用いるのが好ましい。

このようにして選抜したクローンにおける本発明遺伝子の塩基配列の決定及び確認は、通常公知の方法を用いて行うことができる。例えば、マキサムーギルバート法（Maxam-Gilbert, Methods Enzymol., 65:499, 1980)やM13 ファージを用いるジデオキシヌクレオチド鎖終結法(Messing, J. et al., Gene, 19:269, 1982）等により行うことができる。

なお、Δ９位不飽和化酵素が実際に発現しているか否かの確認は例えば、和田らの方法（J. Bacteriol., 175:6056, 1993 ）に従って行うことができる。

上記のようにして塩基配列が決定された本発明遺伝子は、通常公知の手段、例えばホスファイト法を用いた市販のＤＮＡシンセサイザーで合成することも可能

である。

本発明遺伝子又は本発明遺伝子の一部を含みΔ9位不飽和化活性を有するポリペプチドをコードするポリヌクレオチドを上記クローンから分離し、これを植物体への遺伝子導入用ベクターに組み込み、このベクターを植物細胞へ導入し、Δ9位不飽和化酵素を植物体中で発現させることにより、所望の植物に低温耐性を付与することができる。

なお、上記の遺伝子導入が可能な植物の種類には特に制限はない。

ここでいう遺伝子導入用ベクターは、Δ9位不飽和化酵素遺伝子が植物体中で安定に発現しうるように構成されることが必要である。具体的には、プロモーター、翻訳調節領域をコードするＤＮＡ鎖、葉緑体への転移ペプチドをコードするＤＮＡ鎖、本発明遺伝子又は本発明遺伝子の一部を含みΔ9位不飽和化活性を有するポリペプチドをコードするポリヌクレオチド、翻訳終止コドンをコードするＤＮＡ鎖及びターミネーターが適切な位置関係で組み込まれていることが必要である。なお、本発明遺伝子以外の遺伝子導入用ベクターの構成要素としては通常公知のものを用いることができる。上記葉緑体への転移ペプチドをコードするＤＮＡ鎖としては、例えばエンドウのリブロース－1，5－二リン酸カルボキシラーゼの小サブユニット遺伝子を当該転移ペプチドをコードするＤＮＡ鎖として好適に用いることができる。プロモーターとしては、例えばカリフラワーモザイクウイルスの35Ｓプロモーターを、またターミネーターとしては、例えばノパリン合成酵素のターミネーターを用いることができる。

植物細胞への遺伝子導入方法としては、通常公知の方法、例えば「"Plant genetic transformation and gene expression; a laboratory manual", Draper, J. et al. eds., Blackwell Scientific Publications, 1988」記載の方法を用いて行うことができる。その例としては、生物的方法であるウィルスを用いる方法、アグロバクテリウムを用いる方法など、物理・化学的方法であるエレクトロポレーション法、ポリエチレングリコール法、マイクロインジェクションなどが挙げられる。これらのうち、タバコを初めとする双子葉植物に対しては、安定な形質転換を確実に行える点から、アグロバクテリウムを用いる方法が好ましい。アグロバクテリウムを用いる方法には、野生型腫瘍プラスミドを用いる中間ベクター

法(Nature, 287 (1980), p.654; Cell, 32 (1983) p.1033; EMBO J., 3 (1984) P.1525)、T－DNA上の腫瘍形成遺伝子領域を欠損させたベクターを利用する中間ベクター法(EMBO J., 2 (1983) P.2143; Bio/Technology, 3 (1985) p.629)、バイナリーベクター法(Bio/Technology, 1 (1983) p.262; Nature, 303 (1983) p.179; Nucl. Acids Res., 12 (1984) p.8711)などがあり、これらのいずれの方法を用いてもよい。アグロバクテリウムを植物に感染させる方法としては、培養細胞への直接接種法、プロトプラスト共存培養法、リーフディスク法等が挙げられるが、直接かつ容易に多数の形質転換植物体を作成することができるという点から、リーフディスク法を使用することが好ましい。

さらに、植物体を再分化させるには、MS-HF培地等の公知の培地に選択用の抗生物質や植物生長ホルモン等を添加した培地で培養すればよい。発根した幼植物体を土壌に移植して栽培すれば、完全な植物体にまで成長させることができる。

完全な植物体にまで成長させた形質転換植物が低温耐性を有しているか否かについては、以下のようにして検討することができる。

低温傷害を受けない温度（例えば２５℃）で検定植物を栽培した後、一時的に（例えば一週間）低温下（例えば４℃）で栽培し、植物への傷害、例えば葉のクロロシスや稔性の低下を測定すること、あるいは、低温下での生長量を対照植物と比較することにより検討できる。

以下実施例をあげて本発明を詳細に説明するが、本発明はこれらの実施例によって限定されるものではない。

〔実施例１〕Anabaena variabilisのΔ12位不飽和化酵素遺伝子（desA）の上流に隣接してあるオープンリーディングフレームのＤＮＡ断片のクローニング

Anabaena variabilis ＩＡＭ Ｍ－３（東京大学分子細胞生物学研究所より分譲）を、約100mlのＢＧ－11培地（"Plant Molecular Biology", Shaw, C.H. ed., p.279, IRL PRESS, 1988）で培養した。25℃、1,000luxの蛍光灯下で毎分120回振とうし、充分菌を生育させた。培養液を室温で5,000gで10分間遠心分離することにより菌体を沈殿物として回収した。

ゲノムＤＮＡを調製するため、菌体を50mlのＡ液（50mM Tris-HCl, 1mM EDTA, pH8.0）に懸濁して洗浄し、遠心分離することにより菌体を沈殿物として回収した。次に、15mlのＢ液（50mM Tris-HCl, 20mM EDTA, 50mM NaCl, 0.25 M sucrose, pH8.0）に懸濁し、Ｂ液で溶解した40mgのリゾチーム（Sigma）を加え37℃で１時間振とうした。次にプロテナーゼＫを15mgとＳＤＳを終濃度で１％になるように加え37℃で１晩振とうした。その後、$NaClO_4$を終濃度で１Ｍになるように加え、さらに20mlのクロロホルム／イソアミルアルコール（24:1）を加えて10分間振とうした後、遠心分離により水層を回収した。クロロホルム／イソアミルアルコール（24: 1)により再抽出した後、水層に50mlのエタノールを加え、ゲノムＤＮＡ調製物をガラス棒に巻き付けて回収した。このＤＮＡ調製物を20mlのＡ液に溶かし、NaClを終濃度で0.1Mにし、さらにRNaseを終濃度で50mg/mlになるように加え、37℃で１時間インキュベートした。次に、Ａ液で飽和した等量のフェノールで２回抽出した後、水層中のゲノムＤＮＡをエタノールを加えることにより沈殿物として回収し、70％エタノールで洗浄後、１mlのＡ液に溶かしAnabaena variabilis のゲノムＤＮＡ溶液とした。

坂本らはAnabaena variabilis 由来の膜脂質に結合した脂肪酸のΔ12位不飽和化酵素遺伝子のクローニングについて発表（1993年日本植物生理学会年会、講演要旨集、No. 3aF04）した際、Δ12位不飽和化酵素遺伝子の上流に隣接してオープンリーディングフレーム（ＯＲＦ）が存在し、これが不飽和化酵素と何らかの関係を有する可能性を報告したが、その機能は同定されていなかった。本発明者らはそのＯＲＦおよび機能に関心をもち、そのＯＲＦのＤＮＡ鎖中の３箇所の塩基配列に着目して、４本のプライマー（配列番号５～配列番号８）を合成し、Anabaena variabilisのゲノムＤＮＡを鋳型としてＰＣＲを行なった。

上記４本のプライマーのうち、配列番号５と６に示された塩基配列を有するプライマーがセンス鎖、配列番号７と８に示された塩基配列を有するプライマーがアンチセンス鎖をコードし、配列番号６と７に示された塩基配列は同一のアミノ酸配列に由来している。センス鎖およびアンチセンス鎖からそれぞれ任意に１種類ずつのプライマーを選び、計４種類のプライマーの組み合わせでＰＣＲを行なった。反応は、100μlの反応液中にプライマーを各20μM、Anabaena

variabilisのゲノムＤＮＡを１μg入れ、GeneAmp PCR Kit（宝酒造）を用いて行なった。反応の温度制御は、95℃（１分）、45℃（１分）、72℃（２分）を１サイクルとして35サイクル行なった。但し、１サイクル目の95℃は３分間とした。反応終了後、反応液10μlを２％アガロースゲルで電気泳動して合成されたＤＮＡを分離し分析した。その結果、配列番号６と８に示された塩基配列を有するプライマーの組み合わせで合成されたＤＮＡ中に、予想される大きさ（約190bp）のＤＮＡ断片が主要なバンドとして検出された。このＤＮＡ（以下、des 9 var という）断片の両末端をKlenowフラグメントで平滑化した後、プラスミドpTZ18R（Pharmacia)のSmaＩ部位にクローニングし、蛍光ＤＮＡシーケンサー（Applied Biosystems）を用いて塩基配列を決定した。得られた塩基配列を配列番号１に示す。この塩基配列から推定されるアミノ酸配列（配列番号２）は、マウスのステアロイル-CoA不飽和化酵素と有意な相同性を示した〔第１図：des 9 var 断片がコードするアミノ酸配列とマウスのステアロイル－ＣｏＡ不飽和化酵素（ＭＳＣＤ２）のアミノ酸配列の比較を示す〕。

次に、des 9 var 断片をプローブとして、Anacystis nidulansのゲノムＤＮＡをサザン分析した。制限酵素XhoＩ，PstＩおよびBamHＩの各々を単独で用いて約0.1μg のAnacystis nidulansのゲノムＤＮＡを切断し、0.8％アガロースゲル電気泳動でＤＮＡ断片を分離後、ナイロンメンブレン（Hybond-N$^+$; Amersham）にブロッティングした。プローブＤＮＡはMultiprime DNA labelling Kit(Amersham）を用いて〔α－^{32}P〕ｄＣＴＰで標識した。６×SSPE[１×SSPEは10mMリン酸緩衝液（pH7.0），１mM EDTA,0.15M NaCl],0.2％ＳＤＳおよび100μg/mlニシン精子ＤＮＡから成る液中で55℃、16時間インキュベーションしてプローブＤＮＡとメンブレンを反応させた。その後、メンブレンを２×SSC〔１×SSCは0.15M NaCl,15mMクエン酸ナトリウム〕中で室温、15分を２回、次いで0.1×SSC中で40℃、15分を２回振とうして洗い、オートラジオグラフィーを行なった。その結果、いずれの制限酵素で切断した場合も１本のＤＮＡ断片のみが検出された（第２図：図中、Ｎｏｎはゲノム DＮＡを制限酵素で切断していないことを示す）。

〔実施例２〕des 9 var 断片と相同性の高いAnacystis nidulansゲノム中のＤＮ

Ａ鎖のクローニング

Anacystis nidulans R2-SPc （東京大学分子細胞生物学研究所より分譲）の培養およびゲノムＤＮＡの調製は、Anabaena variabilis の場合と同様に行なった。約100μgのゲノムＤＮＡをSau3AIで部分消化した後、Molecular Cloning 2nd edition, pp. 2. 85-2. 87(Sambrook, J. et al. eds., Cold Spring Harbor Laboratory, 1989）の方法に従って、ショ糖密度勾配下での超遠心分離により約９から23kbpのＤＮＡ断片を回収した。これをBamHＩとHindIIIで切断したラムダファージベクターλDASH II(Stratagene）にクローニングした後、ファージ粒子にパッケージングしAnacystis nidulansのゲノムＤＮＡライブラリーを得た。このファージライブラリーを大腸菌P2392 に感染させ、ＮＺＹＭ培地を入れた直径約15cmのシャーレにまいて総数約10万個のプラークを形成させた後、ナイロンメンブレン （Hybond-N$^+$; Amersham）にブロッティングした。上記のサザン分析と同様に、〔$\alpha-{}^{32}$Ｐ〕ｄＣＴＰで標識したdes 9 var 断片をこのメンブレンと反応させ、オートラジオグラフィーによって検出した陽性ファージを再度同様にスクリーニングすることにより、シグナル強度の異なる約30個のファージクローンを得た。この中から任意に12クローンを選び、常法に従ってファージＤＮＡを得た。得られたファージＤＮＡを数種類の制限酵素で切断し、0.8％アガロースゲル電気泳動で分離後、ナイロンメンブレンにブロッティングした。このメンブレンを上記のスクリーニングと同じ条件でサザン分析し、プローブＤＮＡとハイブリダイズするＤＮＡ断片の長さとそのシグナル強度を比較した。その結果、λ５とλ15の２クローンが最も強いシグナルを示し、またインサートＤＮＡ断片の長さもそれぞれ11および15kbpであったため目的のＯＲＦ全体を含むのに十分と判断し、この２クローンのインサートＤＮＡにつき更に幾つかの制限酵素で切断してサザン分析を行なった。その結果XhoIで切断しハイブリダイズすると、２クローンとも約５kbpのＤＮＡ断片が検出されたので、これをpBluescript SK-(Stratagene）のXhoＩサイトにサブクローニングし、λ５とλ15由来のＤＮＡ断片をそれぞれ含むプラスミドｐ５Ｘとｐ15Ｘを得た。ｐ５Ｘとｐ15Ｘの詳細な制限酵素地図を作り比較したところ、ともに同一のゲノムＤＮＡ断片を含むと判断された〔第３図：λ５、λ15およびｐ15ＸのインサートＤＮＡ断片の相互関係

を示す。網かけした長方形はスクリーニングの過程でプローブのdes 9 var 断片がハイブリダイズしたＤＮＡ断片を示す。太い矢印はdes 9 nid（後述）の領域とセンス鎖の方向を示す。細い矢印はdes 9 nid を含む領域のシーケンスの方向を示す。5, 1. 25および0. 5kbpの各バーは左の各図におけるサイズマーカーを示す。制限酵素の略号は、B, BamHI ; H, HindIII ; N, NotI ; Hp, HpaI ; RI, EcoRI ; RV, EcoRV ; S, SalI ; P, PstI ; X, XhoIを示す〕。

そこで、制限酵素あるいはExoIIIによるディリーションプラスミドをｐ15Ｘより作成し、des 9 var 断片がハイブリダイズする領域を含む約２kbpのＤＮＡ断片の塩基配列を蛍光ＤＮＡシーケンサーを用いて決定した（第３図）。その結果そのＤＮＡ断片中には834bpからなるＯＲＦ（des 9 nid）が存在し（配列番号３）、278残基のアミノ酸がコードされていると推定された（配列番号４）。先にクローニングしたAnabaena variabilis 由来のdes 9 var 断片がコードしているアミノ酸配列（配列番号２）との相同性は約80％であった。さらに、核酸およびアミノ酸配列の解析ソフト（ＧＥＮＥＴＹＸ；ソフトウエア開発）と核酸およびアミノ酸配列のデータベース（ＥＭＢＬおよびＤＤＢＪ）を用いて相同性の高いアミノ酸配列の検索を行なったところ、マウスのステアロイル－ＣｏＡ不飽和化酵素との相同性が全体では約３０％であるが局所的に非常に高いこと〔第４図：des 9 nidとマウスのステアロイル－ＣｏＡ不飽和化酵素（ＭＳＣＤ２）のアミノ酸配列の比較〕から、取得したdes 9 nidは脂肪酸を不飽和化する酵素をコードすることが強く示唆された。

〔実施例３〕　des９nid遺伝子の大腸菌での発現による活性測定

Anacystis nidulansは不飽和化酵素として脂質に結合した飽和脂肪酸の９位を不飽和化するΔ９位不飽和化酵素活性しか持たない（Bishop, D. G. et al., Plant Cell Physiol., 27:1593, 1986）ため、des９nidがコードするポリペプチドを大腸菌で発現させ活性測定することを試みた。

p15Xから直接発現させることは困難なので、大腸菌での発現用のベクターを作成した。即ち、ベクターとしてpET3a(Novagen)用い、そのNdeIとBamHIの間にdes９nidをアミノ末端に余計なアミノ酸を付けない用にしてクローニングすることを以下のようにして行なった。des９nidのコードするタンパクのＣ末端側直後にBamHIサイトを入れるために、Ｃ末端を鋏む２箇所の塩基配列を使ってＰＣＲ反応を行なった。即ち、

センスプライマー；5'-ACGTCATGGCCTGCAGT（下線はPstIサイト）（配列番号９）

アンチセンスプライマー；5'-CGCGGATCCTTAGTTGTTTGGAGACG（１重線はBamHIサイト、２重線はストップコドン）（配列番号１０）

p15Xを鋳型として上記の２つのプライマーを用いてＰＣＲ反応を行なうと約１４０ｂｐの産物が得られ、これをpUC19のSmaI部位にサブクローニングして塩基配列に間違いのないことを確認した。この結果得られたプラスミドのBamHIの下流にはEcoRI部位が生じた。これをEcoRIとPstIで順に切断し、一方、p15Xも同じ制限酵素で切断することにより、ストップコドンの直後にBamHI部位を導入した。このプラスミドをSalIで切断した後、４種のdNTP存在下でDNAポリメラーゼKlenow断片を用いてFill in反応を行ない、引き続きHindIIIで切断した。これに、以下の２種の合成DNAから成るアダプターを導入する事によりアミノ末端側にNdeI部位を導入した。即ち、

5'-CATATGACCCTTGCTATCCGACCCA（下線はNdeI）（配列番号１１）及び

5'-AGCTTGGGTCGGATAGCAAGGGTCATATG（１重線はNdeIサイト、２重線はHindIIIの一部）（配列番号１２）

を等モル量混合しアダプターとした。以上のようにして出来たプラスミド（pDes9Nde）を、常法（Molecular cloning pp. 250-251; 1982）に従って調整した大腸菌株BL21(DE3)（Novagen)のコンピテントセルに導入し、アンピシリン耐性

による選別により形質転換株ＢＬＤＥＳ１を得た。

ＢＬＤＥＳ１及びpET3aのみを有するBL21株（ＢＬ１）を100mlのM9培地（200 μg/mlのアンピシリン，4 mg/ml グルコース、10μM FeCl3, 0.5μg/mlビタミンB1, 1 mg/mlカザミノ酸を含む）に接種し、３７℃で培養した。培養液の濁度が、波長600nmで0.5O.D.になるまで培養を続けた後、イソプロピルチオガラクトシド(IPTG)を最終濃度１mMになるように加えた。更に１時間培養し、Δ９位不飽和化酵素遺伝子の発現を誘導した。回収した大腸菌ペレットを1.2% NaClで洗った後、脂質を抽出した。脂質はBlighとDyerの方法（Can J. Biochem. Physiol., 37: 911, 1959）に従って抽出し、2.5 mlの５％塩酸メタノールで完全密封化して85℃２時間半反応させ脂肪酸をメチル化した。生じた脂肪酸メチルエステルを2.5 mlのヘキサンで４回抽出し、窒素ガスで溶媒を除去して濃縮した。脂肪酸メチルエステルの分析には、ガスクロマトグラフィーを用いた。脂肪酸の同定は標準脂肪酸メチルとの保持時間を比較して行なった。定量にはクロマトパックC-R7A plus（島津製作所）を用いた。結果を次の第１表に示す。

第１表.　大腸菌の脂肪酸組成

菌株名	１６：０	１６：１	１８：１（１１）	その他
ＢＬ１（０時間）	４７	２０	２９	４
ＢＬ１（１時間）	５０	１７	２９	４
ＢＬＤＥＳ１（０時間）	４４	２２	３０	４
ＢＬＤＥＳ１（１時間）	４０	２８	２８	４

ここで時間はIPTGによるタンパクの誘導時間を示す。

ＢＬＤＥＳ１では１６：１が増加していることが明らかである。即ち、本遺伝子は１６：０への不飽和化活性を有することが示された。

また、これらの菌株を0.1 mMのステアリン酸を含むM9培地で培養し、同様に比較したところ、ＢＬ１株に比べＢＬＤＥＳ１では１６：１のみならず、１８：１（９）も生成し、des 9 nid がコードするポリペプチドは１６：０ばかりでなく１８：０も基質として不飽和脂肪酸を作出することが示された。

〔実施例４〕 des９nid遺伝子のタバコ植物体への導入

Anacystis nidulans由来のdes９nid遺伝子を次のようにしてタバコに組み込んだ。

（１）植物発現用ベクタープラスミドの構築

pDes9NdeをSacIとSalIで切断する事により両酵素の切断部位で挟まれたdes９nid遺伝子断片が得られる。一方、エンドウのRuBisCO遺伝子を含むクローンpSNIP9（Schreicherら、EMBO J. 4,25(1985)）から葉緑体へのtransit配列をHindIIIとSphIで切り出し、それと同一の制限酵素で切断したpUC118にクローニングすることにより、transit配列の下流にマルチクローニングサイトを有するプラスミド（pTRA3）を得た。このHindIIIサイトを切断後Klenow酵素でFill inしXbaIリンカーをいれた（pTRA3X）。このプラスミドpTRA3XをSal IとSac Iで切断し、さきに同一の制限酵素で切断する事によって得たdes９nid遺伝子断片を挿入した（pTRA3Xdes9）。このプラスミドではRuBisCOのtransit配列に引き続き、それと同一の読み枠でdes９nid遺伝子が翻訳される。これをSac IとXba Iで切断して次に述べる植物用のベクターに挿入する。植物発現型バイナリープラスミドpBI121(Clonetech)を制限酵素SacIとXbaIで切断して得たプラスミドpBI(-GUS)はβ-Glucuronidase遺伝子（GUS遺伝子）を含んでおらず、これにカリフラワーモザイクウイルスの35Sプロモーターとノパリン合成酵素（NOS）ターミネーターの間に前述した導入遺伝子を挿入することにより、植物への導入用ベクター（pBI121(-GUS)Rbsc-des9）を得た。

（２）pBI121(-GUS)Rbsc-des9のアグロバクテリウムへの導入

Agrobacterium tumefaciens LBA4404（Clonetech)を50mlのYEB培地（１l当たりビーフエキス５g、酵母エキス１g、ペプトン１g、ショ糖５g、2mM MgSO4(pH7.4)）に接種し、28℃で24時間培養後、培養液を3,000rpm、４℃、20分の遠心で集菌した。菌体を10mlの１ mM Hepes-KOH(pH7.4)で３回洗った後、３mlの10%グリセロールで１回洗い、最終的に３mlの10%グリセロールに懸濁してＤＮＡ導入用アグロバクテリウムとした。

このようにして得た菌液50μl及び前記のプラスミドpBI121(-GUS)Rbsc-des9 １μgをキュベットに入れ、エレクトロポレーション装置（Gene Pulser;

BioRad）を用いて 25μF、 2500V、 200Ωの条件で電気パルスをかけ、プラスミドＤＮＡをアグロバクテリウムに導入した。この菌液をエッペンドルフチューブに移し、800μlの SOC培地（１ l当たりトリプトン 20 g、酵母エキス５ g、 NaCl 0.5 g、2.5 mM KCl、10 mM MgSO4，10 mM MgCl2，20 mM グルコース、pH7.0）を加え、28℃で 1.5時間静置培養した。この培養液 50μlを、 100 ppmのカナマイシンを含む YEB寒天培地（寒天 1.2%）上にまき、 28℃で２日間培養した。

得られたコロニー群からシングルコロニーを選び、このコロニーからアルカリ法でプラスミドＤＮＡを調整した。このプラスミドＤＮＡを適当な制限酵素で消化後、１％アガロースゲル電気泳動によりＤＮＡ断片を分離し、32Ｐでラベルしたdes９nid遺伝子断片をプローブとしたサザン分析により、プラスミドpBI121(-GUS)Rbsc-des9を含んでいることを確認した。このAgrobacterium tumefaciensをALBBSDESと呼ぶ。

（３）タバコの形質転換

上記の菌株ALBBSDESを、50ppmのカナマイシンを含むLB液体培地で28℃、２時間振とう培養した。培養液1.5 mlを 10,000rpm、３分間遠心して集菌後、カナマイシンを除くために１mlのLB培地で洗浄した。更に10,000rpm、３分間遠心して集菌後、1.5 mlのLB培地に再懸濁し感染用菌液とした。

タバコへの感染に当たっては、若い葉を採取し、0.5%次亜塩素酸ナトリウム水溶液に10分間浸せき後、滅菌水で３回洗い、滅菌済みの濾紙上で水を拭って感染用の葉とした。この葉を１片が１cm² になるようにメスで無菌的に切断し、上記のアグロバクテリウムの菌液上に葉の裏を上にして置き、２分間静かに振とうした後、滅菌済みの濾紙上に葉を置いて過剰のアグロバクテリウムを除いた。シャーレ内のMS-B5培地（ベンジルアデニン1.0 ppm、ナフタレン酢酸 0.1 ppm、及び寒天 0.8 %を含む）（ Murashige, T. and Skoog, F. Plant Physiol., 15: 473, (1962)）上に、ワットマン No. 1濾紙（φ 7.0 cm）を置き、この濾紙に裏を上にして葉を置いた。シャーレをパラフィルムでシールし、 16時間明、８時間暗の周期で 25℃、２日間培養した。ついでクラフォラン 250 ppmを含むMS-B5培地上に移し、同様に 10日間培養してアグロバクテリウムを除去した。

更にクラフォラン 250 ppm及びカナマイシン 100 ppmを含む MS-B5培地上に置床し、同様に７日間培養した。この間に葉片の周囲がカルス化し、シュート原基が生じた。更に 10日間培養後、伸張したシュートをクラフォラン 250 ppm及びカナマイシン 100 ppmを含む MS-HF培地（ベンジルアデニン及びナフタレン酢酸を含まない MS-B5培地）に置床した。 10日間培養後、発根したシュートをカナマイシン耐性の形質転換体とし、プラントボックス内のクラフォラン 250 ppmを含む MS-HF培地に移植した。

〔実施例５〕形質転換タバコのゲノムサザン及びノーザン分析

目的遺伝子の導入を確認するため、カナマイシン耐性のタバコからＤＮＡを抽出し、サザン及びノーザン分析を行った。ゲノムＤＮＡの抽出法はＣＴＡＢ法で成書（Rogers, S. O. & Bendich, A. J.; Plant Molecular Biology Manual A6; 1(1988)）に従って行なった。即ち、２ｇのタバコの葉を液体窒素内で粉砕し、ＣＴＡＢ抽出緩衝液でゲノムＤＮＡを得た。１０μｇのＤＮＡを制限酵素EcoRIとXbaIで切断後0.7%アガロースゲルで電気泳動し、その後ナイロン膜（Hybond N+; Amersham）に0.4 N NaOHで転写した。この膜にpTRA3Xdes9からtransit付きの不飽和化酵素遺伝子をプローブとして、、６５℃で１６時間ハイブリダイゼーションすることにより目的遺伝子がタバコゲノムに組み込まれていることを確認した。

また、導入遺伝子の発現を調べるために、タバコの葉約２ｇからＲＮＡの分析を行なった。方法はグアニジウムチオシアン酸による抽出を行ない（Nagy, F. ら: Plant Molecular Biology Manual B4; 1 (1988)）、poly(A)+RNAをホルムアルデヒド入りのアガロースゲルで電気泳動後、ナイロン膜（Hybond N; Amersham）に転写し、サザン法と同様のハイブリダイゼーションにより分析した。様々の量のＲＮＡを発現している個体があったが、その中から発現量の多い個体について脂質分析を行なった。

〔実施例６〕形質転換タバコの脂質の脂肪酸分析

実施例５でＲＮＡの高発現が確認されたタバコ形質転換体、及び対照としてpBI121で形質転換したタバコの葉から、以下の方法によりホスファチジルグリセロール（ＰＧ）、スルフォキノボシルジアシルグリセロール（ＳＱＤＧ）等の脂

質を調整し、その脂肪酸組成を分析した。なお、一部の個体からは根の脂質も分析した。

（１）全脂質の抽出

脂質の抽出はBligh-Dyer法（Can J. Biochem. Physiol., 37: 911, 1959）で行なった。湿重量２ｇの葉（一部の根を試料とするときは１ｇ）をメスで細断し、これに20 mlのクロロホルム：メタノール（１：２、体積比）を加え、ホモジナイザーで葉を破砕後、 15分間静置した。これにクロロホルム 12ml及び蒸留水12ml を加え激しく混合した後、3000rpm 、４℃、30分間の遠心で水層と有機層の２層に分け、有機層（下層）を回収した。これに適当量のエタノールを加えて、ロータリーエバポレーターを用い、30℃減圧下で溶媒を除いた。これを２ mlのクロロホルム：メタノール（１：４、体積比）に溶かし、全脂質抽出物とした。この一部を５％メタノール性塩酸を用いて後述の方法により処理することで、メチル化脂肪酸を得た。

（２）脂質の分画

DEAE-Toyopearl 650C（東ソー）の懸濁液2.5mlを１ M酢酸ナトリウム水溶液(pH7.0)25 mlと混ぜ酢酸型とした。これを、蒸留水、メタノールで順次洗浄し、最後にメタノールに懸濁して、内径２ cmのカラムに高さ1.5cmまで詰め、更に50 mlのクロロホルム：メタノール（１：４、体積比）で洗浄した。

次に、全脂質抽出物をカラムにかけ、50 mlのクロロホルム：メタノール（１：４、体積比）でモノガラクトシルジアシルグリセロール（ＭＧＤＧ）、ジガラクトシルジアシルグリセロール（ＤＧＤＧ）、ホスファチジルエタノールアミン（ＰＥ）、ホスファチジルコリン（ＰＣ）を溶出して、中性脂質（ＭＧＤＧ、ＤＧＤＧ、ＰＥ、ＰＣ）画分とした。次に５mlの酢酸でホスファチジルセリン（ＰＳ）を溶出して除き、20mlのクロロホルム：メタノール（１：４、体積比）で酢酸を洗浄した後、50 mlのクロロホルム：メタノール：10 M酢酸アンモニウム水溶液（20:80:0.2、体積比）でＰＧ、ＳＱＤＧ、ホスファチジルイノシトール（ＰＩ）を含む画分を得た。この画分に 15mlのエタノールを加え、減圧下で溶媒を除いた。これを0.2mlのクロロホルム：メタノール（２：１、体積比）に溶かし、酸性脂質（ＰＧ、ＳＱＤＧ、ＰＩ）画分とした。

ＭＧＤＧ、ＤＧＤＧ、ＰＥ、ＰＣ画分は、ケイ酸カラムクロマトグラフィー（イアトロビーズ、ヤトロン社）により、さらに分画した。即ち、クロロホルム１mlに溶かした試料をクロロホルムで平衡化したカラムにかけ、クロロホルム：アセトン（４：１）、アセトン、メタノールで順に溶出すると、糖脂質（ＭＧＤＧ、ＤＧＤＧ）はアセトンで、リン脂質（ＰＣ、ＰＥ）はメタノールで溶出された。

（３）薄層クロマトグラフィー（TLC）によるＰＧの単離精製と脂肪酸分析

（２）で得た画分をシリカゲルーTLCプレート#5721（Merck）で分離した。展開溶媒としては、酸性脂質の場合はクロロホルム：アセトン：メタノール：酢酸：水（50:20:10:15:5、体積比）を、中性脂質の場合はクロロホルム：メタノール：水(70:21:3、体積比)を用いた。TLCで分離後、プリムリン（８０％アセトン溶液）を噴霧して紫外線光下で蛍光発色させ、標準となる脂質と移動度を比較することにより各クラスの脂質の画分を推定し、発色した脂質をシリカゲルごと削り取りネジ栓付試験管に入れた。この脂肪酸組成を推定する場合には、この試験管に３ mlのメタノール性５％塩酸を加え、完全密封下85℃で２時間半反応させ、脂肪酸メチル化した。一方、ｓｎ－１、２ごとの脂肪酸組成を決めるために、削り取ったシリカゲルから５ｍｌのクロロホルム：メタノール（２：１）混液で脂質を回収し、乾固した後、１ mlの50 mM TrisCl (pH 7.2)及び0.05 % Triton X-100 を加え、激しく攪拌して脂質を分散させて、クモノスカビ（Rhizopus delemar）由来のリパーゼ（２５００Ｕ；ベーリンガー社）を加え３７℃で３０分間保温することにより選択的にｓｎ－１位の脂肪酸を分解させた。この反応産物を濃縮後、ＴＬＣ（クロロホルム：アセトン：メタノール：酢酸：水＝１０：４：２：３：１）により未反応の脂質、リゾ体、及び脂肪酸に分離した。これらもゲルから回収し前述のようにメタノール性塩酸でメチル化脂肪酸を得た。生じた脂肪酸メチルエステルを３ mlのヘキサンで４回抽出し、減圧下で溶媒を除去して濃縮した。脂肪酸メチルの分析には、ガスクロマトグラフィーを用いた。脂肪酸の同定は標準脂肪酸メチルとの保持時間を比較して行なった。定量にはクロマトパックC-R7A plus（島津製作所）を用いた。全脂質の結果を第２表、PGについて第３表、その他の代表的な脂質ごとの分析結果を第４表に示す。表は、対

照植物体については2個体、形質転換体については独立した2又は3個体の分析値の平均値を示している。

第2表. 葉の全脂質の脂肪酸分析結果

植物	16:0	16:1	16:2	16:3	18:0	18:1	18:2	18:3	Σ16:0+18:0
対照植物体	17	3	1	4	3	1	9	63	20
形質転換体	10	12	1	5	1	2	14	56	11

第3表. ＰＧの脂肪酸分析結果

	植物	16:0	16:1t	16:1c	18:0	18:1	18:2	18:3	Σ16:0+18:0+16:1t
PG	対照植物体	32	37	0	1	5	10	14	70
	形質転換体	18	37	8	0	10	12	15	55

第４表．その他の脂質の脂肪酸分析結果

	植物	16:0	16:1	16:2	16:3	18:0	18:1	18:2	18:3	Σ16:0+18:0
SQDG	対照植物体	51	1	0	0	3	2	7	36	54
	形質転換体	36	22	0	0	0	4	9	28	36
MGDG	対照植物体	7	0	1	9	1	2	4	76	8
	形質転換体	3	9	1	10	0	1	5	69	3
DGDG	対照植物体	19	0	0	0	3	1	4	73	22
	形質転換体	9	13	0	1	0	1	5	70	9
PC	対照植物体	28	0	0	0	5	1	21	44	33
	形質転換体	19	12	0	0	3	4	40	23	22
PE	対照植物体	20	0	0	0	3	1	6	70	23
	形質転換体	18	10	0	0	2	2	31	38	20
PI	対照植物体	48	1	0	0	2	1	11	37	50
	形質転換体	44	7	0	0	1	2	18	28	45

ＰＧに結合した脂肪酸分析の結果から、Anacystis nidulans由来の脂肪酸不飽和化酵素を発現している形質転換タバコでは１６：０（パルミチン酸）が大幅に減り、そのかわりに１６：１ｃｉｓが増えている事、また、少量在った１８：０もほとんど無くなり、逆に１８：１が増えているいることが判明した。その結果、飽和脂肪酸（１６：０＋１６：１ｔｒａｎｓ＋１８：０（ステアリン酸））含量は、対照のタバコでは７０％であるのに対し、不飽和化酵素の遺伝子を形質転換したタバコでは５５％と著しく低くなっている。ＰＧのｓｎ－１、２位別の分析結果から、ｓｎ－２は９８％以上飽和脂肪酸（１６：０又は１６：１ｔｒａｎｓ）で占められており、新たに遺伝子導入により生成した１６：１はすべてｓｎ－１にあることが判明した。従って、この不飽和化酵素遺伝子を形質転換したタバコのＰＧのｓｎ－１位の飽和脂肪酸は極めて少なくなっていることが明らかである。従って、ｓｎ－１、２両位共に、飽和脂肪酸から成る、所謂、飽和分子種の量も大幅に減少し、脂質の分子種の組成上著しく低温に耐性な型に変化した事がわかる。

一方、その他の脂質のＭＧＤＧ、ＤＧＤＧ、ＳＱＤＧ、ＰＣ、ＰＥ、ＰＩでも、１６：０の減少と、それに呼応した１６：１の１０％前後の増加が明らかであり、また、１８：０の不飽和化も進んでいた。このうち、ＭＧＤＧとＤＧＤＧについては１６：１の生成は主としてｓｎ－１位にあったが、ｓｎ－２位からも少量検出された。ＭＧＤＧ、ＤＧＤＧ、ＳＱＤＧ及びＰＧは主に葉緑体に存在する脂質であり、ラン藻であるAnacystis nidulansの不飽和化酵素を高等植物の葉緑体で発現させることにより驚くほど不飽和化が進展したことがわかる。それにも増して、これらの４種の脂質はAnacystis nidulansの膜にも存在する物であり、不飽和化の基質になる可能性は高かったが、それ以外のＰＣ、ＰＥ及びＰＩはAnacystis nidulansの膜には存在しない脂質であり、しかも高等植物では主に葉緑体外に存在する脂質であることから、それらにおいてもパルミチン酸、及びステアリン酸が不飽和化されたのは驚くべきことである。

このように、形質転換タバコの脂質分析の結果から、Anacystis nidulans由来の脂肪酸不飽和化酵素が、高等植物であるタバコの形質転換体において、ほとんど総ての脂質の１６：０と１８：０を極めて効率良く不飽和化できることを本実施例は証明した。

また、根の全脂質について、脂肪酸分析をした結果を第５表に示す。

第５表．根の全脂質の脂肪酸分析

植物	16:0	16:1	16:2	16:3	18:0	18:1	18:2	18:3	Σ16:0+18:0
非形質転換体	26	0	0	0	5	2	47	21	31
形質転換体	17	13	0	0	2	4	58	6	19

この結果から、Anacystis nidulans由来の脂肪酸不飽和化酵素は、驚くべきことに、葉のみならず根においても１６：０と１８：０の不飽和化を触媒したこと

がわかる。このことは、本発明の脂肪酸不飽和化酵素遺伝子が植物の低温耐性を変化させる可能性ばかりでなく、不飽和脂肪酸含量を増やす可能性を有し、植物を油の原料とする産業においても有用であることを示す。

〔実施例７〕形質転換タバコの低温耐性試験

上記のＲＮＡの発現解析及び脂質分析で有望と思われた個体については、自殖することにより次世代の種子を採取した。その一部をカナマイシン 800 ppmを含む MS-HF培地に蒔き、２５℃、１６時間明、８時間暗の日長で２週間栽培後にカナマイシン耐性の実生を選抜した。この実生をプラントボックスに移植して、更に４週間栽培した。また、コントロールとして、pBI121により形質転換した個体についても、上記の操作を行った。

次に、４℃連続光のもとで１１日間低温処理した後、２５℃で２日間栽培した。その結果、コントロール植物（pBI121により形質転換した植物）では葉に対して大幅な萎縮症状及びクロロシスが観察されたのに対し、形質転換植物ではほとんど傷害は観察されなかった。従って、不飽和化酵素遺伝子の導入により低温耐性が向上したと推定された。

産業上の利用可能性

本発明のΔ９位不飽和化酵素をコードする遺伝子を植物に導入することにより、植物に低温耐性を付与することおよび植物中の不飽和脂肪酸含量を増やすことが可能となった。

【配列表】

配列番号：1

配列の長さ：１９６

配列の型：核酸

鎖の数：二本鎖

トポロジー：直鎖状

配列の種類：Genomic DNA

起源

生物名：Anabaena variabilis

株名： IAM M-3

配列：

GCT CTG GGG TTG TTG CTG TTA TAT CTA GGC GGG TGG TCT TTT GTG GTC TGG
GGA GTT TTC TTT CGC ATC GTT TGG GTT TAC CAC TGT ACT TGG TTG GTA AAC
AGC GCT ACC CAT AAG TTT GGC TAC CGC ACC TAT GAT GCT GGT GAC AGA TCC
ACT AAC TGT TGG TGG GTA GCT GTC CTA GTG TTT GGT GAA GGT T

配列番号：2

配列の長さ：６５

配列の型：アミノ酸

トポロジー：直鎖状

配列の種類：ペプチド

起源

生物名：Anabaena variabilis

株名： IAM M-3

配列：

Ala Leu Gly Leu Leu Leu Leu Tyr Leu Gly Gly Trp Ser Phe Val Val Trp
Gly Val Phe Phe Arg Ile Val Trp Val Tyr His Cys Thr Trp Leu Val Asn Ser
Ala Thr His Lys Phe Gly Tyr Arg Thr Tyr Asp Ala Gly Asp Arg Ser Thr Asn
Cys Trp Trp Val Ala Val Leu Val Phe Gly Glu Gly

配列番号：3

配列の長さ：８３７

配列の型：核酸

鎖の数：二本鎖

トポロジー：直鎖状

配列の種類：Genomic DNA

起源

生物名：Anacystis nidulans

株名：R2-SPc

配列：

ATG ACC CTT GCT ATC CGA CCC AAG CTT GCC TTC AAC TGG CCG ACC GCC CTG
TTC ATG GTC GCC ATT CAC ATT GGA GCA CTG TTA GCG TTC CTG CCG GCC AAC
TTT AAC TGG CCC GCT GTG GGC GTG ATG GTT GCG CTG TAT TAC ATT ACC GGT
TGT TTT GGC ATC ACC CTA GGC TGG CAC CGG CTA ATT TCG CAC CGT AGC TTT
GAA GTT CCC AAA TGG CTG GAA TAC GTG CTG GTG TTC TGT GGC ACC TTG GCC
ATG CAG CAC GGC CCG ATC GAA TGG ATC GGT CTG CAC CGC CAC CAT CAC CTC
CAC TCT GAC CAA GAT GTC GAT CAC CAC GAC TCC AAC AAG GGT TTC CTC TGG
AGT CAC TTC CTG TGG ATG ATC TAC GAA ATT CCG GCC CGT ACG GAA GTA GAC
AAG TTC ACG CGC GAT ATC GCT GGC GAC CCT GTC TAT CGC TTC TTT AAC AAA
TAT TTC TTC GGT GTC CAA GTC CTA CTG GGG GTA CTT TTG TAC GCC TGG GGC
GAG GCT TGG GTT GGC AAT GGC TGG TCT TTC GTC GTT TGG GGG ATC TTC GCC
CGC TTG GTG GTG GTC TAC CAC GTC ACT TGG CTG GTG AAC AGT GCT ACC CAC
AAG TTT GGC TAC CGC TCC CAT GAG TCT GGC GAC CAG TCC ACC AAC TGC TGG
TGG GTT GCC CTT CTG GCC TTT GGT GAA GGC TGG CAC AAC AAC CAC CAC GCC
TAC CAG TAC TCG GCA CGT CAT GGC CTG CAG TGG TGG GAA TTT GAC TTG ACT
TGG TTG ATC ATC TGC GGC CTG AAG AAG GTG GGT CTG GCT CGC AAG ATC AAA
GTG GCG TCT CCA AAC AAC TAA

配列番号：4

配列の長さ：２７８

配列の型：アミノ酸

トポロジー：直鎖状

配列の種類：ペプチド

起源

生物名：Anacystis nidulans

株名： R2-SPc

配列：

Met Thr Leu Ala Ile Arg Pro Lys Leu Ala Phe Asn Trp Pro Thr Ala Leu Phe
Met Val Ala Ile His Ile Gly Ala Leu Leu Ala Phe Leu Pro Ala Asn Phe Asn
Trp Pro Ala Val Gly Val Met Val Ala Leu Tyr Tyr Ile Thr Gly Cys Phe Gly
Ile Thr Leu Gly Trp His Arg Leu Ile Ser His Arg Ser Phe Glu Val Pro Lys
Trp Leu Glu Tyr Val Leu Val Phe Cys Gly Thr Leu Ala Met Gln His Gly Pro
Ile Glu Trp Ile Gly Leu His Arg His His His Leu His Ser Asp Gln Asp Val
Asp His His Asp Ser Asn Lys Gly Phe Leu Trp Ser His Phe Leu Trp Met Ile
Tyr Glu Ile Pro Ala Arg Thr Glu Val Asp Lys Phe Thr Arg Asp Ile Ala Gly
Asp Pro Val Tyr Arg Phe Phe Asn Lys Tyr Phe Phe Gly Val Gln Val Leu Leu
Gly Val Leu Leu Tyr Ala Trp Gly Glu Ala Trp Val Gly Asn Gly Trp Ser Phe
Val Val Trp Gly Ile Phe Ala Arg Leu Val Val Val Tyr His Val Thr Trp Leu
Val Asn Ser Ala Thr His Lys Phe Gly Tyr Arg Ser His Glu Ser Gly Asp Gln
Ser Thr Asn Cys Trp Trp Val Ala Leu Leu Ala Phe Gly Glu Gly Trp His Asn
Asn His His Ala Tyr Gln Tyr Ser Ala Arg His Gly Leu Gln Trp Trp Glu Phe
Asp Leu Thr Trp Leu Ile Ile Cys Gly Leu Lys Lys Val Gly Leu Ala Arg Lys
Ile Lys Val Ala Ser Pro Asn Asn

配列番号：5

配列の長さ：１８

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：他の核酸　合成DNA

配列：

ATGACAATTG CTACTTCA

配列番号：６

配列の長さ：１５

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：他の核酸　合成DNA

配列：

GCTCTGGGGT TGTTG

配列番号：７

配列の長さ：１５

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：他の核酸　合成DNA

配列：

CAACAACCCC AGAGC

配列番号：８

配列の長さ：１８

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：他の核酸　合成DNA

配列：

RTGRTGRTTR TTRTGCCA

配列番号：９

配列の長さ：１７

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：他の核酸　合成DNA

配列：

ACGTCATGGC CTGCAGT

配列番号：１０

配列の長さ：２６

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：他の核酸　合成DNA

配列：

CGCGGATCCT TAGTTGTTTG GAGACG

配列番号：１１

配列の長さ：２５

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：他の核酸　合成DNA

配列：

CATATGACCC TTGCTATCCG ACCCA

配列番号：１２

配列の長さ：２９

配列の型：核酸

鎖の数：一本鎖

トポロジー：直鎖状

配列の種類：他の核酸　合成DNA

配列：

AGCTTGGGTC GGATAGCAAG GGTCATATG

請 求 の 範 囲

1. 脂質に結合した脂肪酸のΔ9位を不飽和化する活性を有するタンパク質をコードする遺伝子。

2. 脂質に結合した脂肪酸のΔ9位を不飽和化する活性を有するタンパク質が実質的に配列番号4に記載されたアミノ酸配列を有するものである請求の範囲第1項記載の遺伝子又は当該遺伝子の一部を含むポリヌクレオチド。

3. 脂質に結合した脂肪酸のΔ9位を不飽和化する活性を有するタンパク質をコードする遺伝子が配列番号3に記載の塩基配列を含むＤＮＡ鎖である請求の範囲第1項記載の遺伝子又は当該遺伝子の一部を含むポリヌクレオチド。

4. 請求の範囲第1項乃至第3項のいずれかの項記載の遺伝子又は当該遺伝子の一部を含むポリヌクレオチドを含むベクター。

5. 請求の範囲第1項乃至第3項のいずれかの項記載の遺伝子又は当該遺伝子の一部を含むポリヌクレオチドが導入された植物細胞。

6. 請求の範囲第5項記載の植物細胞を分化させて植物体を再生させることを特徴とする植物の作出方法。

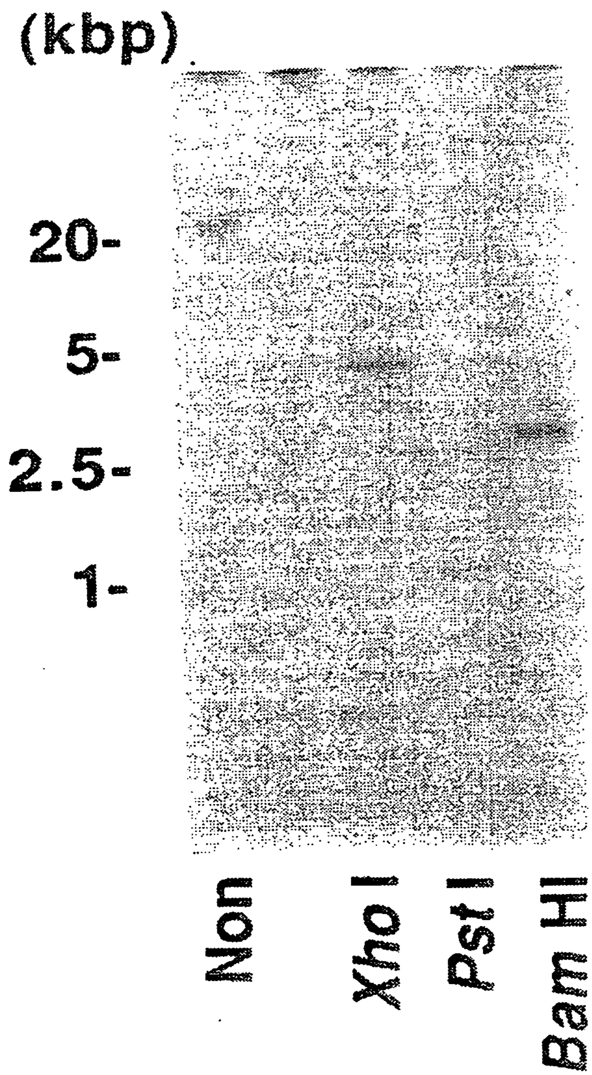
7. 請求の範囲第1項乃至第3項のいずれかの項記載の遺伝子又は当該遺伝子の一部を含むポリヌクレオチドが導入された植物。

第１図

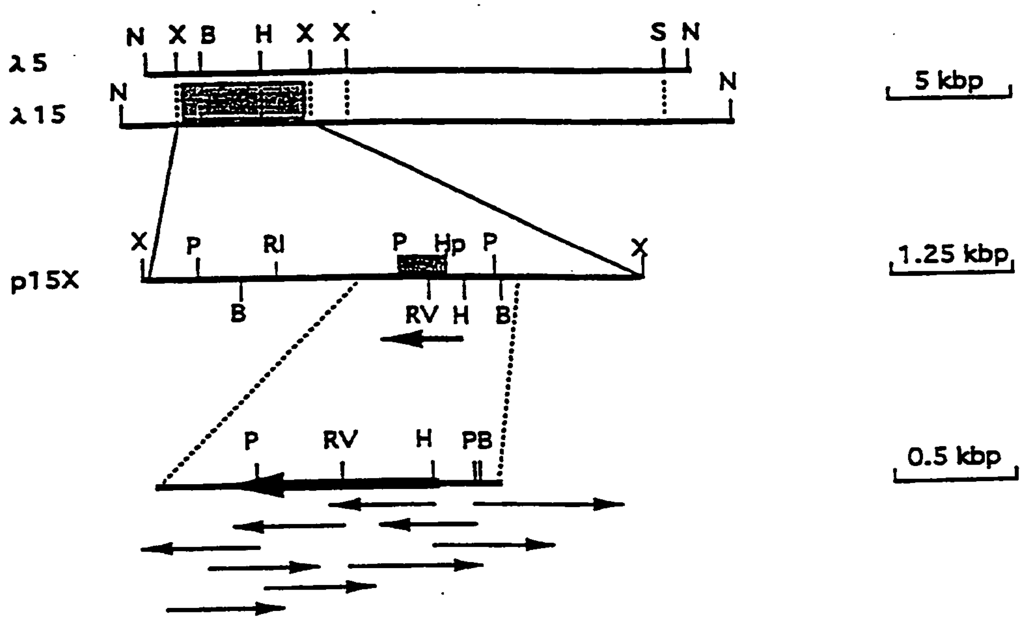
```
                   10         20        30        40        50        60
des9var  ALGLLLLYLGGWSFVVWGVFFRIVWVYHCTWLVNSATHKFGYRTYDAGDRSTNCWWVAVL
                          ..:.: . :.  X:::::::.: .:::.::     :     :
MSCD2    LVPWYCWGETFVNSLCVSTFLRYAVVLNATWLVNSAAHLYGYRPYDKNISSRENILVSMG
                  240       250       260       270       280       290

des9var  VFGEG
           :X
MSCD2    AVGER
```

第2図



第３図


λ5
λ15
N
X
B
H
X
X
S
N
N
N
5 kbp
p15X
X
P
RI
P
Hp
P
X
B
RV
H
B
1.25 kbp
P
RV
H
PB
0.5 kbp

第4図

```
                         10        20        30        40        50        60
des9nid     MTLAIRPKLAFNWPTALFMVAIHIGALLAFLPANFNWPAVGVMVALYYITGCFGITLGWH
                  X::.. :.. ..:. .:.::: ..  .      . ... :::... .::: : :
MSCD2       DDEGPPPKLEYVWRNIILMALLHLGALYGITLVPSCKLYTCLFAYLYYVISALGITAGAH
           60        70        80        90       100       110

                  70        80        90       100       110       120
des9nid     RLISHRSFEVPKWLEYVLVFCGTLAMQHGPIEWIGLHRHHHLHSDQDVDHHDSNKGFLWS
            :: :::.....  :   :.. .:.:.:..  ::   :: ::  :. ..:.:.:..::..:
MSCD2       RLWSHRTYKARLPLRLFLIIANTMAFQNDVYEWARDHRAHHKFSETHADPHNSRRGFFFS
           120       130       140       150       160       170

                  130        140       150       160       170
des9nid     HFLW-MIYEIPA-RTEVDKFTRDIAGDPVYRFFNKYFFGVQVLLGVLLYAWGEAWVGNGW
            X  : .. . :: ... .:.....  .    .:.. ..  ..::  .. .    .:   :
MSCD2       HVGWLLVRKHPAVKEKGGKLDMSDLKAEKLVMFQRRYYKPDLLLMCFVLPTLVPWYCWGE
           180       190       200       210       220       230

             180         190       200       210       220       230
MSCD2       SFV---VWGIFARLVVVYHVTWLVNSATHKFGYRSHESGDQSTNCWWVALLAFGEGWHNN
            .::    ..: : .:: ..:::::::.: .:::.....   : .    :.. : :: .::
MSCD2       TFVNSLCVSTFLRYAVVLNATWLVNSAAHLYGYRPYDKNISSRENILVSMGAVGERFHNY
           240       250       260       270       280       290

              240       250       260       270
des9nid     HHAYQYSARHGLQWWEFDLTWLIICGLKKVGLARKIKVASPNN
            :::..:. . .  .:....: ..:  .  .:::
MSCD2       HHAFPYDYSASEYRWHINFTTFFIDCMALLGLAYDRKRVSRAA
           300       310       320       330       340
```

第5図

不飽和化酵素遺伝子　　対照

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP94/02288

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl⁶ C12N15/00, C12N5/00, A01H1/00, A01H5/00//C12N9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int. Cl⁶ C12N15/00, C12N5/00, A01H1/00, A01H5/00//C12N9/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CAS ONLINE
WPI, WPI/L, BIOSIS PREVIEWS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X P, Y	Journal of Biological Chemistry, Vol. 269, No. 41 (1994) T. Sakamoto, et al. "Delta, 9 Acyl-lipid desaturases of cyanobacteria. Molecular cloning and substrate specificities in terms of fatty acids, sn-positions, and polar head groups" p. 25576-25580	1, 4 5-7
X Y	EP, A, 561569 (Lubrizol Corp.), September 22, 1993 (22. 09. 93) & AU, A, 9335167 & CA, A, 2092661 & JP, A, 6014667	1, 4-7 5-7
X Y	NL, A, 9002130 (Stichting Tech Wetenschappen), April 16, 1992 (16. 04. 92)	1, 4-7 5-7
X Y	Journal of Biological Chemistry, Vol. 265, No. 33 (1990) J. E. Stukey, et al "The OLEI gene of Saccharomyces cerevisiae encodes the delta-9 fatty acid desaturase and can be furctionally replaced by the rat stearoylcoA desaturase gene" p. 20144-20149	1, 4 5-7

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
March 14, 1995 (14. 03. 95)	April 4, 1995 (04. 04. 95)

Name and mailing address of the ISA/	Authorized officer
Japanese Patent Office	
Facsimile No.	Telephone No.

国際調査報告

国際出願番号 PCT／JP 94／02288

A. 発明の属する分野の分類（国際特許分類（IPC））

Int. Cl⁶ C12N15／00，C12N5／00，A01H1／00，A01H5／00 // C12N9／00

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int. Cl⁶ C12N15／00，C12N5／00，A01H1／00，A01H5／00 // C12N9／00

最小限資料以外の資料で調査を行った分野に含まれるもの

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

CAS ONLINE，
WPI，WPI／L，BIOSIS PREVIEWS

C. 関連すると認められる文献

引用文献のカテゴリー＊	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
P，X P，Y	Journal of Biological chemistry. 第269巻，第41号（1994）T. Sakamoto, et al.「Delta,9 Acyl－lipid desaturases of cyanobacteria. Molecular clonirg and substrate specificities in terms of fatty acids，sn－positions，and polar head groups」p. 25576－25580	1，4 5－7

☑ C欄の続きにも文献が列挙されている。 ☐ パテントファミリーに関する別紙を参照。

＊ 引用文献のカテゴリー
「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」先行文献ではあるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
14. 03. 95	04.04.95

名称及びあて先
日本国特許庁（ISA／JP）
郵便番号100
東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員） 4B 9281
權野浩志 ㊞
電話番号 03－3581－1101 内線 3449

国際調査報告 | 国際出願番号 PCT/JP 94/02288

C（続き）． 関連すると認められる文献		
引用文献のカテゴリー＊	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
X / Y	EP, A, 561569(Lubrizol Corp.), 22. 9月. 1993(22. 09. 93) &AU, A, 9335167&CA, A, 2092661 &JP, A, 6014667	1.4−7 / 5−7
X / Y	NL, A, 9002130 (Stichting Tech Wetenschappen), 16. 4月. 1992(16. 04. 92)	1.4−7 / 5−7
X / Y	Journal of Biological Chemistry，第265巻，第33号 (1990) J. E. Stukey, et al「The OLEI gene of Saccharomyces cerevisiae encodes the delta−9 fatty acid desaturase and can be furctionally replaced by the rat stearoylcoA desaturase gene」p. 20144−20149	1.4 / 5−7

(1)

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 1/21, 15/29, 15/82
C07H 15/12

A1

(11) International Publication Number: WO 91/13972

(43) International Publication Date: 19 September 1991 (19.09.91)

(21) International Application Number: PCT/US91/01746

(22) International Filing Date: 14 March 1991 (14.03.91)

(30) Priority data:

494,106	16 March 1990 (16.03.90)	US
567,373	13 August 1990 (13.08.90)	US
615,784	14 November 1990 (14.11.90)	US

(60) Parent Application or Grant
(63) Related by Continuation
US 494,106 (CIP)
Filed on 16 March 1990 (16.03.90)

(71) Applicant *(for all designated States except US):* CALGENE, INC. [US/US]; 1920 Fifth Street, Davis, CA 95616 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : THOMPSON, Gregory, A. [US/US]; 5127 Cowell Blvd., Davis, CA 95616 (US). KNAUF, Vic, C. [US/US]; 1013 Hillview Lane, Winters, CA 95694 (US).

(74) Agents: LASSEN, Elizabeth et al.; Calgene, Inc., 1920 Fifth Street, Davis, CA 95616 (US).

(81) Designated States: AT (European patent), BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PLANT DESATURASES - COMPOSITIONS AND USES

(57) Abstract

By this invention, compositions and methods of use of plant desaturase enzymes, especially Δ-9 desaturases, are provided. Of special interest are methods and compositions of amino acids and nucleic acid sequences related to biologically active plant desaturases as well as sequences, especially nucleic acid sequences, which are to be used as probes, vectors for transformation or cloning intermediates. Biologically active sequences may be found in a sense or anti-sense orientation as to transcriptional regulatory regions found in various constructs.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

PLANT DESATURASES - COMPOSITIONS AND USES

This application is a continuation-in-part of USSN 07/494,106 filed on March 16, 1990 and a continuation-in-part of USSN 07/567,373 filed on August 13, 1990 and a continuation-in-part of USSN 07/615,784 filed on November 14, 1990.

Technical Field

The present invention is directed to desaturase enzymes relevant to fatty acid synthesis in plants, enzymes, amino acid and nucleic acid sequences and methods related thereto, and novel plant entities and/or oils and methods related thereto.

INTRODUCTION

Background

Novel vegetable oils compositions and/or improved means to obtain or manipulate fatty acid compositions, from biosynthetic or natural plant sources, are needed. Depending upon the intended oil use, various different oil compositions are desired. For example, edible oil sources containing the minimum possible amounts of saturated fatty acids are desired for dietary reasons and alternatives to current sources of highly saturated oil products, such as tropical oils, are also needed.

One means postulated to obtain such oils and/or modified fatty acid compositions is through the genetic engineering of plants. However, in order to genetically engineer plants one must have in place the means to transfer genetic material to the plant in a stable and heritable manner. Additionally, one must have nucleic acid sequences capable of producing the desired phenotypic result, regulatory regions capable of directing the correct application of such sequences, and the like. Moreover, it

should be appreciated that to produce a desired modified oils phenotype requires that the Fatty Acid Synthetase (FAS) pathway of the plant is modified to the extent that the ratios of reactants are modulated or changed.

Higher plants appear to synthesize fatty acids via a common metabolic pathway in plant plastid organelles (i.e., chloroplasts, proplastids, or other related organelles) as part of the FAS complex. Outside of plastid organelles, fátty acids are incorporated into triglycerides and used in plant membranes and in neutral lipids. In developing seeds, where oils are produced and stored as sources of energy for future use, FAS occurs in proplastids.

The production of fatty acids begins in the plastid with the reaction between Acyl Carrier Protein (ACP) and acetyl-CoA to produce acetyl-ACP. Through a sequence of cylical reactions, the acetyl-ACP is elongated to 16- and 18- carbon fatty acids. The longest chain fatty acids produced by the FAS are 18 carbons long. Monunsaturated fatty acids are also produced in the plastid through the action of a desaturase enzyme.

Common plant fatty acids, such as oleic, linoleic and α-linolenic acids, are the result of sequential desaturation of stearate. The first desaturation step is the desaturation of stearoyl-ACP (C18:0) to form oleoyl-ACP (C18:1) in a reaction often catalyzed by a Δ-9 desaturase, also often referred to as a "stearoyl-ACP desaturase" because of its high activity toward stearate the 18 carbon acyl-ACP. The desaturase enzyme functions to add a double bond at the ninth carbon in accordance with the following reaction (I):

$$\text{Stearoyl-ACP} + \text{ferredoxin(II)} + O_2 + 2H^+ \rightarrow \text{oleoyl-ACP} + \text{ferredoxin (III)} + 2H_2O.$$

Δ-9 desaturases have been studied in partially purified preparations from numerous plant species. Reports indicate that the protein is a dimer, perhaps a homodimer, displaying a molecular weight of 68 kD (±8 kD) by gel-filtration and a molecular weight of 36 kD by SDS-polyacrylamide gel electrophoresis.

In subsequent sequential steps for triglyceride production, polyunsaturated fatty acids may be produced. These desaturations occur outside of the plastid as a result of the action of membrane-bound enzymes. Additional double bonds are added at the twelve position carbon and thereafter, if added, at the 15 position carbon through the action of Δ-12 desaturase and Δ-15 desaturase, respectively.

Obtaining nucleic acid sequences capable of producing a phenotypic result in FAS, desaturation and/or incorporation of fatty acids into a glycerol backbone to produce an oil is subject to various obstacles including but not limited to the identification of metabolic factors of interest, choice and characterization of a protein source with useful kinetic properties, purification of the protein of interest to a level which will allow for its amino acid sequencing, utilizing amino acid sequence data to obtain a nucleic acid sequence capable of use as a probe to retrieve the desired DNA sequence, and the preparation of constructs, transformation and analysis of the resulting plants.

Thus, the identification of enzyme targets and useful plant sources for nucleic acid sequences of such enzyme targets capable of modifying fatty acid compositions are needed. Ideally, an enzyme target will be amenable to one or more applications alone or in combination with other nucleic acid sequences relating to increased/decreased oil production, the ratio of saturated to unsaturated fatty acids in the fatty acid pool, and/or to novel oils compositions as a result of the modifications to the fatty acid pool. Once enzyme target(s) are identified and qualified, quantities of protein and purification protocols are needed for sequencing. Ultimately, useful nucleic acid constructs having the necessary elements to provide a phenotypic modification and plants containing such constructs are needed.

Relevant Literature

A 200-fold purification of *Carthamus tinctorius* ("safflower") stearoyl-ACP desaturase was reported by McKeon & Stumpf in 1982, following the first publication of their protocol in 1981. McKeon, T. & Stumpf, P. *J.Biol.Chem.* (1982) *257*:12141-12147; McKeon, T. & Stumpf, P. *Methods in Enzymol.* (1981) *71*:275-281.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 provides amino acid sequence of fragments relating to *C. tinctorius* desaturase. Fragments F1 through F11 are also provided in the sequence listing as SEQ ID NO: 1 through SEQ ID NO: 11, respectively. Each fragment represents a synthesis of sequence information from peptides originating from different digests which have been matched and aligned. In positions where there are two amino acids indicated, the top one corresponds to that found in the translation of the cDNA; the lower one was detected either as a second signal at the same position of one of the sequenced peptides, or as a single unambiguous signal found in one or more of the overlapping peptides comprising the fragment. Residues in F9 shown in lower case letters represent positions where the called sequence does not agree with that predicted from the cDNA, but where the amino acid assignment is tentative because of the presence of a contaminating peptide. The standard one letter code for amino acid residues has been used. X represents a position where no signal was detectable, and which could be a modified residue. F1 corresponds to the N-terminal sequence of the mature protein. The underlined region in F2 is the sequence used in designing PCR primers for probe synthesis.

Fig. 2 provides a cDNA sequence (SEQ ID NO: 12) and the corresponding translational peptide sequence (SEQ ID NO: 13) derived from *C. tinctorius* desaturase. The cDNA sequence includes both the plastid transit peptide encoding sequence and the sequence encoding the mature protein.

Fig. 3 provides cDNA sequence of *Ricinus communis* desaturase. Fig. 3A provides preliminary partial cDNA sequence of a 1.7 kb clone of *R. communis* desaturase (SEQ ID NO: 14). The sequence is from the 5' end of the clone. Fig. 3B provides the complete cDNA sequence of the approximately 1.7 kb clone (SEQ ID NO: 15) and the corresponding translational peptide sequence (SEQ ID NO: 16).

Fig. 4 provides sequence of *Brassica campestris* desaturase. Fig. 4A represents partial DNA sequence of a 1.6 kb clone pCGN3235 (SEQ ID NO: 17), from the 5' end of the clone. Fig. 4B represents partial DNA sequence of a 1.2 kb clone, pCGN3236, from the 5' end of the clone (SEQ ID NO: 18). Initial sequence for the 3' ends of the two *B. campestris* desaturase clones indicates that pCGN3236 is a shorter cDNA for the same clone as pCGN3235. Fig. 4C provides complete cDNA sequence of *B. campestris* desaturase above, pCGN3235 (SEQ ID NO: 19) and the corresponding translational peptide sequence (SEQ ID NO: 20).

Fig. 5 provides preliminary partial cDNA sequence of *Simmondsia chinensis* desaturase (SEQ ID NO: 43). The translated amino acid sequence is also shown.

Fig. 6 shows the design of forward and reverse primers (SEQ ID NO: 21 through SEQ ID NO: 26) used in polymerase chain reaction (PCR) from the sequence of *C. tinctorius* desaturase peptide "Fragment F2" (SEQ ID NO: 2).

Fig. 7 provides maps of desaturase cDNA clones showing selected restriction enzyme sites. Fig. 7A represents a *C. tinctorius* clone, Fig. 7B represents a *R. communis* clone, and Fig. 7C represents a *B. campestris* clone.

Fig. 8 provides approximately 3.4 kb of genomic sequence of Bce4 (SEQ ID NO: 27).

Fig. 9 provides approximately 4 kb of genomic sequence of Bcg 4-4 ACP sequence (SEQ ID NO: 28).

Fig. 10 provides a restriction map of cloned λCGN 1-2 showing the entire napin coding region sequence as well as extensive 5' upstream and 3' downstream sequences (SEQ ID NO: 29).

SUMMARY OF THE INVENTION

By this invention, compositions and methods of use of plant desaturase enzymes, especially Δ-9 desaturases, are provided. Of special interest are methods and compositions of amino acids and nucleic acid sequences related to biologically active plant desaturases as well as sequences, especially nucleic acid sequences, which are to be used as probes, vectors for transformation or cloning intermediates. Biologically active sequences may be found in a sense or anti-sense orientation as to transcriptional regulatory regions found in various constructs.

A first aspect of this invention relates to C. tinctorius Δ-9 desaturase substantially free of seed storage protein. Amino acid sequence of this desaturase is provided in Fig. 2 and as SEQ ID NO: 13.

DNA sequence of *C. tinctorius* desaturase gene (SEQ ID NO: 12) is provided, as well as DNA sequences of desaturase genes from a *Ricinus* (SEQ ID NO: 14 and SEQ ID NO: 15) a *Brassica* (SEQ ID NO: 17 through SEQ ID NO: 19) and a *Simmondsia* (SEQ ID NO: 43) plant.

In yet a different embodiment of this invention, plant desaturase cDNA of at least 10 nucleotides or preferably at least 20 nucleotides and more preferably still at least 50 nucleotides, known or homologously related to known Δ-9 desaturase(s) is also provided. The cDNA encoding precursor desaturase or, alternatively, biologically active, mature desaturase is provided herein.

Methods to use nucleic acid sequences to obtain other plant desaturases are also provided. Thus, a plant desaturase may be obtained by the steps of contacting a nucleic acid sequence probe comprising nucleotides of a known desaturase sequence and recovery of DNA sequences encoding plant desaturase having hybridized with the probe.

This invention also relates to methods for obtaining plant Δ-9 desaturase by contacting an antibody specific to a known desaturase, such as *C. tinctorius* stearoyl-ACP

desaturase, with a candidate plant stearoyl-ACP desaturase under conditions conducive to the formation of an antigen:antibody immunocomplex and the recovery of the candidate plant stearoyl-ACP desaturase which reacts thereto.

In a further aspect of this invention DNA constructs comprising a first DNA sequence encoding a plant desaturase and a second DNA sequence which is not naturally found joined to said plant desaturase are provided. This invention also relates to the presence of such constructs in host cells, especially plant host cells. In yet a different aspect, this invention relates to transgenic host cells which have an expressed desaturase therein.

Constructs of this invention may contain, in the 5' to 3' direction of transcription, a transcription initiation control regulatory region capable of promoting transcription in a host cell and a DNA sequence encoding plant desaturase. Transcription initiation control regulatory regions capable of expression in prokatyotic or eukaryotic host cells are provided. Most preferred are transcription initiation control regions capable of expression in plant cells, and more preferred are transcription and translation initiation regions preferentially expressed in plant cells during the period of lipid accumulation. The DNA sequence encoding plant desaturase of this invention may be found in either the sense or anti-sense orientation to the transcription initiation control region.

Specific constructs, expression cassettes having in the 5' to 3' direction of transcription, a transcription and translation initiation control regulatory region comprising sequence immediately 5' to a structural gene preferentially expressed in plant seed during lipid accumulation, a DNA sequence encoding desaturase, and sequence 3' to the structural gene are also provided. The construct may preferably contain DNA sequences encoding plant desaturase obtainable (included obtained) from *Carthamus*, *Rininus*, *Brassica* or *Simmondsia* Δ-9 desaturase

genes. Transcription and translation initiation control regulatory regions are preferentially obtained from structural genes preferentially expressed in plant embryo tissue such as napin, seed-ACP or Bce-4.

By this invention, methods and constructs to inhibit the production of endogenous desaturase are also provided. For example, an anti-sense construct comprising, in the 5' to 3' direction of transcription, a transcription initiation control regulatory region functional in a plant cell, and an anti-sense DNA sequence encoding a portion of a plant Δ-9 desaturase may be integrated into a plant host cell to decrease desaturase levels.

In yet a different embodiment, this invention is directed to a method of producing plant desaturase in a host cell comprising the steps of growing a host cell comprising an expression cassette, which would contain in the direction of transcription, a) a transcription and translation initiation region functional in said host cell, b) the DNA sequence encoding a plant desaturase in reading frame with said initiation region, and c) and a transcript termination region functional in said host cell, under conditions which will promote the expression of the plant desaturase. Cells containing a plant desaturase as a result of the production of the plant desaturase encoding sequence and also contemplated herein.

By this invention, a method of modifying fatty acid composition in a host plant cell from a given level of fatty acid saturation to a different level of fatty acid saturation is provided by growing a host plant cell having integrated into its genome a recombinant DNA sequence encoding a plant desaturase in either a sense or anti-sense orientation under control of regulatory elements functional in said plant cell during lipid accumulation under conditions which will promote the activity of said regulatory elements. Plant cells having such a modified level of fatty acid saturation are also contemplated hereunder. Oilseeds having such a modified level of fatty

acid saturation and oils produced from such oilseeds are further provided.

DETAILED DESCRIPTION OF THE INVENTION

A plant desaturase of this invention includes any sequence of amino acids, such as a protein, polypeptide, or peptide fragment, obtainable from a plant source which is capable of catalyzing the insertion of a first double bond into a fatty acyl-ACP moiety in a plant host cell, i.e., *in vivo*, or in a plant cell-like environment, i.e. *in vitro*. "A plant cell-like environment" means that any necessary conditions are available in an environment (i.e., such factors as temperatures, pH, lack of inhibiting substances) which will permit the enzyme to function In particular, this invention relates to enzymes which add such a first double bond at the ninth carbon position in a fatty acyl-ACP chain. There may be similar plant desaturase enzymes of this invention with different specificities, such as the Δ-12 desaturase of carrot.

Nucleotide sequences encoding desaturases may be obtained from natural sources or be partially or wholly artificially synthesized. They may directly correspond to a desaturase endogenous to a natural plant source or contain modified amino acid sequences, such as sequences which have been mutated, truncated, increased or the like. Desaturases may be obtained by a variety of methods, including but not limited to, partial or homogenous purification of plant extracts, protein modeling, nucleic acid probes, antibody preparations and sequence comparisons. Typically a plant desaturase will be derived in whole or in part from a natural plant source.

Of special interest are Δ-9 desaturases which are obtainable, including those with are obtained, from *Cartharmus*, *Ricinus*, *Simmondsia*, or *Brassica*, for example *C. tinctorius*, *R. communis*, *S. chinensis* and *B. campestris*, respectively, or from plant desaturases which are obtainable through the use of these sequences. "Obtainable" refers to those desaturases which have

sufficiently similar sequences to that of the native sequences provided herein to provide a biologically active desaturase.

Once a DNA sequence which encodes a desaturase is obtained, it may be employed as a gene of interest in a nucleic acid construct or in probes in accordance with this invention. A desaturase may be produced in host cells for harvest or as a means of effecting a contact between the desaturase and its substrate. Constructs may be designed to produce desaturase in either prokaryotic or eukaryotic cells. Plant cells containing recombinant constructs encoding biologically active desaturase sequences, both expression and anti-sense constructs, as well as plants and cells containing modified levels of desaturase proteins are of special interest. For use in a plant cell, constructs may be designed which will effect an increase or a decrease in amount of endogenous desaturase available to a plant cell transformed with such a construct.

Where the target gene encodes an enzyme, such as a plant desaturase, which is already present in the host plant, there are inherent difficulties in analyzing mRNA, engineered protein or enzyme activity, and modified fatty acid composition or oil content in plant cells, especially in developing seeds; each of which can be evidence of biological activity. This is because the levels of the message, enzyme and various fatty acid species are changing rapidly during the stage where measurements are often made, and thus it can be difficult to discriminate between changes brought about by the presence of the foreign gene and those brought about by natural developmental changes in the seed. Where an expressed Δ-9 desaturase DNA sequence is derived from a plant species heterologous to the plant host into which the sequence is introduced and has a distinguishable DNA sequence, it is often possible to specifically probe for expression of the foreign gene with oligonucleotides complimentary to unique sequences of the inserted DNA/RNA. And, if the foreign gene codes for a protein with slightly different protein sequence, it may be

possible to obtain antibodies which recognize unique epitopes on the engineered protein. Such antibodies can be obtained by mixing the antiserum to the foreign protein with extract from the host plant, or with extracts containing the host plant enzyme. For example, one can isolate antibodies uniquely specific to a *C. tinctorius* Δ-9 desaturase by mixing antiserum to the desaturase with an extract containing a *Brassica* Δ-9 desaturase. Such an approach will allow the detection of *C. tinctorius* desaturase in *Brassica* plants transformed with the *C. tinctorius* desaturase gene. In plants expressing an endogenous gene in an antisense orientation, the problem is slightly different. In this case, there are no specific reagents to measure expression of a foreign protein. However, one is attempting to measure a decrease in an enzyme activity that normally is increasing during development. This makes detection of expression a simpler matter. In the final seed maturation phase, enzyme activities encoded by genes affecting oil composition usually disappear and cannot be detected in final mature seed. Analysis of the fatty acid content may be preformed by any manner known to those skilled in the art, including gas chromatography, for example.

By increasing the amount of desaturase available in the plant cell, an increased percentage of unsaturated fatty acids may be provided; by decreasing the amount of desaturase, an increased percentage of saturated fatty acids may be provided. (Modifications in the pool of fatty acids available for incorporation into triglycerides may likewise affect the composition of oils in the plant cell.) Thus, an increased expression of desaturase in a plant cell may result in increased proportion of fatty acids, such as one or more of palmitoleate (C16:1), oleate (C18:1), linoleate (C18:2) and linolenate (C18:3) are expected. Of special interest is the production of triglycerides having increased levels of oleate. Using anti-sense technology, alternatively, a decrease in the amount of desaturase available to the plant cell is expected, resulting in a

higher percentage of saturates such as one or more of laurate (C12:0), myristate (C14:0), palmitate (C16:0), stearate (C18:0), arachidate (C20:0), behemate (C22:0) and lignocerate (C24:0). Of special interest is the production of triglycerides having increased levels of stearate or palmitate and stearate. In addition, the production of a variety of ranges of such saturates is desired. Thus, plant cells having lower and higher levels of stearate fatty acids are contemplated. For example, fatty acid compositions, including oils, having a 10% level of stearate as well as compositions designed to have up to a 60% level of stearate or other such modified fatty acid(s) composition are contemplated.

The modification of fatty acid compositions may also affect the fluidity of plant membranes. Different lipid concentrations have been observed in cold-hardened plants, for example. By this invention, one may be capable of introducing traits which will lend to chill tolerance. Constitutive or temperature inducible transcription initiation regulatory control regions may have special applications for such uses.

Other applications for use of cells or plants producing desaturase may also be found. For example, potential herbicidal agents selective for plant desaturase may be obtained through screening to ultimately provide environmentally safe herbicide products. The plant desaturase can also be used in conjunction with chloroplast lysates to enhance the production and/or modify the composition of the fatty acids prepared *in vitro*. The desaturase can also be used for studying the mechanism of fatty acid formation in plants and bacteria. For these applications, constitutive promoters may find the best use.

Constructs which contain elements to provide the transcription and translation of a nucleic acid sequence of interest in a host cell are "expression cassettes". Depending upon the host, the regulatory regions will vary, including regions from structural genes from viruses, plasmid or chromosomal genes, or the like. For expression

in prokaryotic or eukaryotic microorganisms, particularly unicellular hosts, a wide variety of constitutive or regulatable promoters may be employed. Among transcriptional initiation regions which have been described are regions from bacterial and yeast hosts, such as *E. coli*, *B. subtilis*, *Saccharomyces cerevisiae*, including genes such as β-galactosidase, T7 polymerase, trp E and the like.

A recombinant construct for expression of desaturase in a plant cell ("expression cassette") will include, in the 5' to 3' direction of transcription, a transcription and translation initiation control regulatory region (the transcriptional and translational initiation regions together often also known as a "promoter") functional in a plant cell, a nucleic acid sequence encoding a plant desaturase, and a transcription termination region. Numerous transcription initiation regions are available which provide for a wide variety of constitutive or regulatable, e.g., inducible, transcription of the desaturase structural gene. Among transcriptional initiation regions used for plants are such regions associated with cauliflower mosaic viruses (35S, 19S), and structural genes such as for nopaline synthase or mannopine synthase or napin and ACP promoters, etc. The transcription/translation initiation regions corresponding to such structural genes are found immediately 5' upstream to the respective start codons. Thus, depending upon the intended use, different promoters may be desired.

Of special interest in this invention are the use of promoters which are capable of preferentially expressing the desaturase in seed tissue, in particular, at early stages of seed oil formation. Examples of such seed-specific promoters include the region immediately 5' upstream of napin or seed ACP genes, such as described in co-pending USSN 147,781, and the Bce-4 gene such as described in co-pending USSN 494,722. Alternatively, the use of the 5' regulatory region associated with an endogenous plant desaturase structural gene and/or the

transcription termination regions found immediately 3' downstream to the gene, may often be desired.

In addition, for some applications, use of more than one promoter may be desired. For example, one may design a dual promoter expression cassette each promoter having a desaturase sequence under its regulatory control. For example, the combination of an ACP and napin cassette could be useful for increased production of desaturase in a seed-specific fashion over a longer period of time than either individually.

To decrease the amount of desaturase found in a plant host cell, anti-sense constructs may be prepared and then inserted into the plant cell. By "anti-sense" is meant a DNA sequence in the 5' to 3' direction of transcription in relation to the transcription initiation region, which encodes a sequence complementary to the sequence of a native desaturase. It is preferred that an anti-sense plant desaturase sequence be complementary to a plant desaturase gene indigenous to the plant host. Sequences found in an anti-sense orientation may be found in constructs providing for transcription or transcription and translation of the DNA sequence encoding the desaturase, including expression cassettes. Constructs having more than one desaturase sequence under the control of more than one promoter or transcription initiation region may also be employed with desaturase constructs. Various transcription initiation regions may be employed. One of ordinary skill in the art can readily determine suitable regulatory regions. Care may be necessary in selecting transcription initiation regions to avoid decreasing desaturase activity in plant cells other than oilseed tissues. Any transcription initiation region capable of directing expression in a plant host which causes initiation of adequate levels of transcription selectively in storage tissues during seed development for example, should be sufficient. As such, seed specific promoters may be desired. Other manners of decreasing the amount of endogenous plant desaturase, such as ribozymes or the

screening of plant cells transformed with constructs for rare events containing sense sequences which in fact act to decrease desaturase expression, are also contemplated herein. Other analogous methods may be applied by those of ordinary skill in the art.

By careful selection of plants, transformants having particular oils profiles may be obtained. This may in part depend upon the qualities of the transcription initiation region(s) employed or may be a result of culling transformation events to exploit the variabilities of expression observed.

In order to obtain the nucleic acid sequences encoding *C. tinctorius* desaturase, a protein preparation free of a major albumin-type contaminant is required. As demonstrated more fully in the Examples, the protocols of McKeon and Stumpf, *supra*, result in a preparation contaminated with a seed storage protein. Removal of the protein contaminant may be effected by application of a reverse-phase HPLC, or alternatively, by application of a reduction and alkylation step followed by electrophoresis and blotting, for example. Other purification methods may be employed as well, now that the presence of the contaminant is confirmed and various properties thereof described. Once the purified desaturase is obtained it may be used to obtain the corresponding amino acid and/or nucleic acid sequences thereto in accordance with methods familiar to those skilled in the art. Approximately 90% of the total amino acid sequence of the *C. tinctorius* desaturase is provided in Fig. 1 and in SEQ ID NOS: 1-11. The desaturase produced in accordance with the subject invention can be used in preparing antibodies for assays for detecting plant desaturase from other sources.

A nucleic acid sequence of this invention may include genomic or cDNA sequence and mRNA. A cDNA sequence may or may not contain pre-processing sequences, such as transit peptide sequences. Transit peptide sequences facilitate the delivery of the protein to a given organelle and are

cleaved from the amino acid moiety upon entry into the organelle, releasing the "mature" sequence.

In Fig. 2 and SEQ ID NO: 13, the sequence of the *C. tinctorius* desaturase precursor protein is provided; both the transit peptide and mature protein sequence are shown. Also provided in this invention are cDNA sequences relating to *R. communis* desaturase (Fig.3 and SEQ ID NOS: 14-15), *B. campestris* desaturase (Fig. 4 and SEQ ID NOS: 17-19) and *S. chinesis* (Fig. 5 and SEQ ID NOS: 43).

The use of the precursor cDNA sequence is preferred in desaturase expression cassettes. In addition, desaturase transit peptide sequences may be employed to translocate other proteins of interest to plastid organelles for a variety of uses, including the modulation of other enzymes related to the FAS pathway. See, European Patent Application Publication No. 189,707.

As described in more detail below, the complete genomic sequence of a desaturase may be obtained by the screening of a genomic library with a desaturase cDNA probe and isolating those sequences which regulate expression in seed tissue. In this manner, the transcription, translation initiation regions and/or transcript termination regions of the desaturase may be obtained for use in a variety of DNA constructs, with or without the respective desaturase structural gene.

Other nucleic acid sequences "homologous" or "related" to DNA sequences encoding other desaturases are also provided. "Homologous" or "related" includes those nucleic acid sequences which are identical or conservatively substituted as compared to the exemplified *C. tinctorius*, *R. communis, S. chinesis* or *B. campestris* desaturase sequences of this invention or a plant desaturase which has in turn been obtained from a plant desaturase of this invention. By conservatively substituted is meant that codon substitutions encode the same amino acid, as a result of the degeneracy of the DNA code, or that a different amino acid having similar properties to the original amino acid is substituted. One skilled in the art will readily

recognize that antibody preparations, nucleic acid probes (DNA and RNA) sequences encoding and the like may be prepared and used to screen and recover desaturase from other plant sources. Typically, nucleic acid probes are labeled to allow detection, preferably with radioactivity although enzymes or other methods may also be used. For immunological screening methods, antibody preparations either monoclonal or polyclonal are utilized. Polyclonal antibodies, although less specific, typically are more useful in gene isolation. For detection, the antibody is labeled using radioactivity or any one of a variety of second antibody/enzyme conjugate systems that are commercially available. Examples of some of the available antibody detection systems are described by Oberfilder (*Focus* (1989) BRL Life Technologies, Inc., *11*:1-5).

A "homologous" or "related" nucleic acid sequence will show at least about 60% homology, and more preferably at least about 70% homology, between the known desaturase sequence and the desired candidate plant desaturase of interest, excluding any deletions which may be present. Homology is determined upon comparison of sequence information, nucleic acid or amino acid, or through hybridization reactions. Amino acid sequences are considered homologous by as little as 25% sequence identity between the two complete mature proteins. (*See generally*, Doolittle, R.F., of URFS and ORFS, University Science Books, CA, 1986.)

Oligonucleotide probes can be considerably shorter than the entire sequence, but should be at least about 10, preferably at least about 15, more preferably at least 20 nucleotides in length. When shorter length regions are used for comparison, a higher degree of sequence identity is required than for longer sequences. Shorter probes are often particularly useful for polymerase chain reactions (PCR), especially when highly conserved sequences can be identified. (*See*, Gould, *et al.*, *PNAS USA* (1989) *86*:1934-1938.) Longer oligonucleotides are also useful, up to the full length of the gene encoding the polypeptide of

interest. When longer nucleic acid fragments are employed (>100 bp) as probes, especially when using compl te or large cDNA sequences, one would screen with low stringencies (for example 40-50°C below the melting temperature of the probe) in order to obtain signal from the target sample with 20-50% deviation, i.e., homologous sequences. (*See*, Beltz, *et al.*, *Methods in Enzymology* (1983) *100*:266-285.) Both DNA and RNA probes can be used.

A genomic library prepared from the plant source of interest may be probed with conserved sequences from a known desaturase to identify homologously related sequences. Use of the entire cDNA may be employed if shorter probe sequences are not identified. Positive clones are then analyzed by restriction enzyme digestion and/or sequencing. When a genomic library is used, one or more sequences may be identified providing both the coding region, as well as the transcriptional regulatory elements of the desaturase gene from such plant source. In this general manner, one or more sequences may be identified providing both the coding region, as well as the transcriptional regulatory elements of the desaturase gene from such plant source.

In use, probes are typically labeled in a detectable manner (for example with ^{32}P-labeled or biotinylated nucleotides) and are incubated with single-stranded DNA or RNA from the plant source in which the gene is sought, although unlabeled oligonucleotides are also useful. Hybridization is detected by means of the label after single-stranded and double-stranded (hybridized) DNA or DNA/RNA have been separated, typically using nitrocellulose paper or nylon membranes. Hybridization techniques suitable for use with oligonucleotides are well known to those skilled in the art. Thus, plant desaturase genes may be isolated by various techniques from any convenient plant. Plant desaturase of developing seed obtained from other oilseed plants, such as soybean, coconut, oilseed rape, sunflower, oil palm, peanut, cocoa, cotton, corn and the like are desired as well as from non-traditional oil

sources, including but not limited to spinach chloroplast, avocado mesocarp, cuphea, California Bay, cucumber, carrot, meadowfoam, *Oenothera* and *Euglena gracillis*.

Once the desired plant desaturase sequence is obtained, it may be manipulated in a variety of ways. Where the sequence involves non-coding flanking regions, the flanking regions may be subjected to resection, mutagenesis, etc. Thus, transitions, transversions, deletions, and insertions may be performed on the naturally occurring sequence. In addition, all or part of the sequence may be synthesized, where one or more codons may be modified to provide for a modified amino acid sequence, or one or more codon mutations may be introduced to provide for a convenient restriction site or other purpose involved with construction or expression. The structural gene may be further modified by employing synthetic adapters, linkers to introduce one or more convenient restriction sites, or the like.

Recombinant constructs containing a nucleic acid sequence encoding a desaturase of this invention may be combined with other, i.e. "heterologous," DNA sequences in a variety of ways. By heterologous DNA sequences is meant any DNA sequence which is not naturally found joined to the native desaturase, including combinations of DNA sequences from the same plant of the plant desaturase which are not naturally found joined together. In a preferred embodiment, the DNA sequence encoding a plant desaturase is combined in a DNA construct having, in the 5' to 3' direction of transcription, a transcription initiation control region capable of promoting transcription in a host cell, and a DNA sequence encoding a desaturase in either a sense or anti-sense orientation. As described in more detail elsewhere, a variety of regulatory control regions containing transcriptional or transcriptional and translational regions may be employed, including all or part of the non-coding regions of the plant desaturase.

The open reading frame coding for the plant desaturase or functional fragment thereof will be joined at its 5' end

to a transcription initiation regulatory control region. In some instances, such as modulation of plant desaturase via a desaturase in an anti-sense orientation, a transcription initiation region or transcription/ translation initiation region may be used. In embodiments wherein the expression of the desaturase protein is desired in a plant host, a transcription/ translation initiation regulatory region, is needed. Additionally, modified promoters, i.e., having transcription initiation regions derived from one gene source and translation initiation regions derived from a different gene source or enhanced promoters, such as double 35S CaMV promoters, may be employed for some applications.

As described above, of particular interest are those 5' upstream non-coding regions which are obtained from genes regulated during seed maturation, particularly those preferentially expressed in plant embryo tissue, such as ACP-and napin-derived transcription initiation control regions. Such regulatory regions are active during lipid accumulation and therefore offer potential for greater control and/or effectiveness to modify the production of plant desaturase and/or modification of the fatty acid composition. Especially of interest are transcription initiation regions which are preferentially expressed in seed tissue, i.e., which are undetectable in other plant parts. For this purpose, the transcript initiation region of acyl carrier protein isolated from *B. campestris* seed and designated as "Bcg 4-4" and an unidentified gene isolated from *B. campestris* seed and designated as "Bce-4" are also of substantial interest.

Briefly, Bce4 is found in immature embryo tissue at least as early as 11 days after anthesis (flowering), peaking about 6 to 8 days later or 17-19 days post-anthesis, and becoming undetectable by 35 days post-anthesis. The timing of expression of the Bce4 gene closely follows that of lipid accumulation in seed tissue. Bce4 is primarily detected in seed embryo tissue and to a lesser extent found in the s ed coat. Bce4 has not been

detected in other plant tissues tested, root, stem and leaves.

Approximately 3.4 kb genomic sequence of Bce4 is provided in Fig. 8 and as SEQ ID NO: 27, including about 1 kb 5' to the structural gene, about 0.3 kb of the Bce4 coding gene sequence, and about 2.1 kb of the non-coding regulatory 3' sequence. Bce4 transcript initiation regions will contain at least 1 kb and more preferably about 5 to about 7.5 kb of sequence immediately 5' to the Bce4 structural gene.

The Bcg 4-4 ACP message presents a similar expression profile to that of Bce4 and, therefore, also corresponds to lipid accumulation in the seed tissue. Bcg 4-4 is not found in the seed coat and may show some differences in expression level, as compared to Bce4, when the Bcg 4-4 5' non-coding sequence is used to regulate transcription or transcription and translation of a plant stearoyl-ACP desaturase of this invention. Genomic sequence of Bcg 4-4 is provided in Fig. 9 and as SEQ ID NO: 28, including about 1.5 kb 5' to the structural gene, about 1.2 kb of the Bcg 4-4 (ACP) structural gene sequence, and about 1.3 kb of the non-coding regulatory 3' sequence.

The napin 1-2 message is found in early seed development and thus, also offers regulatory regions which can offer preferential transcriptional regulation of a desired DNA sequence of interest such as the plant desaturase DNA sequence of this invention during lipid accumulation. Napins are one of the two classes of storage proteins synthesized in developing *Brassica* embryos (Bhatty, *et al., Can J. Biochem.* (1968) *46*:1191-1197) and have been used to direct tissue-specific expression when reintroduced into the *Brassica* genome (Radke, *et al., Theor. Appl. Genet.* (1988) *75*:685-694). Genomic sequence of napin 1-2 is provided in Fig. 10 and as SEQ ID NO: 29, including about 1.7 kb 5' to the structural gene and about 1.3 kb of the non-coding regulatory 3' sequence

Regulatory transcript termination regions may be provided in DNA constructs of this invention as well.

Transcript termination regions may be provided by the DNA sequence encoding the plant desaturase or a convenient transcription termination region derived from a different gene source, especially the transcript termination region which is naturally associated with the transcript initiation region. The transcript termination region will contain at least about 1 kb, preferably about 3 kb of sequence 3' to the structural gene from which the termination region is derived.

In developing the DNA construct, the various components of the construct or fragments thereof will normally be inserted into a convenient cloning vector which is capable of replication in a bacterial host, e.g., *E. coli*. Numerous vectors exist that have been described in the literature. After each cloning, the plasmid may be isolated and subjected to further manipulation, such as restriction, insertion of new fragments, ligation, deletion, insertion, resection, etc., so as to tailor the components of the desired sequence. Once the construct has been completed, it may then be transferred to an appropriate vector for further manipulation in accordance with the manner of transformation of the host cell.

Normally, included with the DNA construct will be a structural gene having the necessary regulatory regions for expression in a host and providing for selection of transformant cells. The gene may provide for resistance to a cytotoxic agent, e.g. antibiotic, heavy metal, toxin, etc., complementation providing prototrophy to an auxotrophic host, viral immunity or the like. Depending upon the number of different host species into which the expression construct or components thereof are introduced, one or more markers may be employed, where different conditions for selection are used for the different hosts.

The manner in which the DNA construct is introduced into the plant host is not critical to this invention. Any method which provides for efficient transformation may be employed. Various methods for plant cell transformation include the use of Ti- or Ri-plasmids, microinjection,

electroporation, liposome fusion, DNA bombardment or the like. In many instances, it will be desirable to have the construct bordered on one or both sides by T-DNA, particularly having the left and right borders, more particularly the right border. This is particularly useful when the construct uses *A. tumefaciens* or *A. rhizogenes* as a mode for transformation, although the T-DNA borders may find use with other modes of transformation.

Where *Agrobacterium* is used for plant cell transformation, a vector may be used which may be introduced into the *Agrobacterium* host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the *Agrobacterium* host. The Ti- or Ri-plasmid containing the T-DNA for recombination may be armed (capable of causing gall formation) or disarmed (incapable of causing gall formation), the latter being permissible, so long as the *vir* genes are present in the transformed *Agrobacterium* host. The armed plasmid can give a mixture of normal plant cell and gall.

A preferred method for the use of *Agrobacterium* as the vehicle for transformation of plant cells employs a vector having a broad host range replication system, at least one T-DNA boundary and the DNA sequence or sequences of interest. Commonly used vectors include pRK2 or derivatives thereof. See, for example, Ditta *et al.*, *PNAS USA*, (1980) 77:7347-7351 and EPA 0 120 515, which are incorporated herein by reference. Normally, the vector will be free of genes coding for opines, oncogenes and *vir*-genes. Included with the expression construct and the T-DNA will be one or more markers, which allow for selection of transformed *Agrobacterium* and transformed plant cells. A number of markers have been developed for use with plant cells, such as resistance to chloramphenicol, the aminoglycoside G418, hygromycin, or the like. The particular marker employed is not essential to this invention, one or another marker being preferred depending on the particular host and the manner of construction.

The expression constructs may be employed with a wide variety of plant life, particularly plant life involved in the production of vegetable oils. These plants include, but are not limited to rapeseed, sunflower, *C. tinctorius*, cotton, *Cuphea*, peanut, soybean, oil palm and corn. Anti-sense constructs may be employed in such plants which share complementarity between the endogenous sequence and the anti-sense desaturase. Of special interest is the use of an anti-sense construct having a *B. campestris* desaturase in rapeseed, including *B. campestris* and *B. napus*.

For transformation of plant cells using *Agrobacterium*, explants may be combined and incubated with the transformed *Agrobacterium* for sufficient time for transformation, the bacteria killed, and the plant cells cultured in an appropriate selective medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of vegetable oils compositions. A variety of stable genetic lines having fixed levels of saturation may be obtained and integrated into a traditional breeding program. Hemizygous and heterozygous lines or homozygous lines may demonstrate different useful properties for oil production and/or breeding. For example, saturation levels may be increased up to 2-fold by the development of homozygous plants as compared with heterozygous (including hemizygous) plants.

The invention now being generally described, it will be more readily understood by reference to the following examples which are included for purposes of illustration only and are not intended to limit the present invention.

EXAMPLES

MATERIALS

Commercially available biological chemicals and chromatographic materials, including BSA, catalase (twice

crystalized from bovine liver), spinach ferredoxin, ferredoxin-$NADP^+$ oxidoreductase (spinach leaf), NADPH, unlabeled fatty acids, DEAE-cellulose (Whatman DE-52) CNBr-activated Sepharose 4B, and octyl-Sepharose, and Reactive Blue Agarose are from Sigma (St. Louis, MO). Triethylamine, trichloroacetic acid, guanidine-HCl, and hydrazine-hydrate are also from Sigma. Proteolytic enzymes, including endoproteinases lysC, gluC, and aspN are sequencing grade enzymes obtained from Boehringer Mannheim (Indianapolis, IN). Organic solvents, including acetone, acetonitrile, methanol, ether and petroleum ether are purchased from J.T. Baker (Phillipsburg, NJ); concentrated acids and sodium sulfate are also from J.T. Baker (Phillipsburg, NJ). HPLC grade acetonitrile and trifluoracetic acid (TFA) are obtained from Burdick and Jackson (Muskegon, MI), and from Applied Biosystems (Foster City, CA), respectively. Radiochemicals, including [9,10(n)-^{3}H] oleic acid (10mCi/μmol) and [^{3}H]-iodoacetic acid (64Ci/mol) are from New England Nuclear (Boston, MA). Phenacyl-8 Reagent (bromoacetophenone with a crown ether catalyst) used to prepare phenacyl esters of the fatty acids for analysis are from Pierce (Rockford, IL). C18 reversed-phase thin-layer chromatography plates are from Whatman (Clifton, NJ).

Acyl carrier protein (ACP) and acyl-ACP synthase are isolated from *E. coli* strain K-12 as described by Rock and Cronan (Rock and Cronan, *Methods in Enzymol* (1981) *71*:341-351 and Rock *et al.*, *Methods in Enzymol.* (1981) *72*:397-403). The *E. coli* is obtainable from Grain Processing (Iowa) as frozen late-logarithmic phase cells.

[9,10(n)-^{3}H]stearic acid is synthesized by reduction of [9,10(n)-^{3}H]oleic acid with hydrazine hydrate essentially as described by Johnson and Gurr (*Lipids* (1971) *6*:78-84). [9,10(n)-^{3}H]oleic acid (2 mCi), supplemented with 5.58mg unlabeled oleic acid to give a final specific radioactivity of 100mCi/mmol, is dissolved in 2ml of acetonitrile, acidified with 40μl of glacial acetic acid, and heated to 55°C. Reduction is initiated with 100μl of

60% (w/w) hydrazine hydrate; oxygen is bubbled through the mixture continuously. After each hour acetonitrile is added to bring the volume back to 2ml and an additional 100µl of hydrazine hydrate is added. At the end of 5 hr. the reaction is stopped by addition of 3ml of 2M HCl. The reaction products are extracted with three 3ml aliquots of petroleum ether and the combined ether extracts are washed with water, dried over sodium sulfate and evaporated to dryness. The dried reaction products are redissolved in 1.0ml acetonitrile and stored at -20°C. The distribution of fatty acid products in a 15µl aliquot is determined by preparation of phenacyl esters, which are then analyzed by thin layer chromatography on C-18 reverse phase plates developed with methanol:water:95:5 (v/v). Usually reduction to [9,10(n)-^{3}H]stearic acid is greater than 90%, a small amount of unreacted oleic acid may remain. The analysis is used to establish fraction of the total radioactivity that is present as stearate, and thereby to determine the exact substrate concentration in the enzyme assay.

Acyl-ACP substrates, including [9,10(n)-^{3}H] stearoyl-ACP are prepared and purified by the enzymatic synthesis procedure of Rock, Garwin, and Cronan (*Methods in Enzymol.* (1981) *72*:397-403).

Acyl carrier protein was covalently bound to Sepharose 4B by reaction of highly purified ACP with CNBr-activated Sepharose 4B as described by McKeon and Stumpf (*J. Biol. Chem.* (1982) *257*:12141-12147).

Example 1

In this example, an initial purification of *C. tinctorius* (safflower) desaturase, following the method of McKeon and Stumpf (*J. Biol. Chem.* (1982) *257*:12141-12142), is described.

Assay: In each of the following steps, the presence of the enzyme is detected radiometrically by measuring enzyme-catalyzed release of tritium from [9,10(n)-

^{3}H]stearoyl-ACP. Preparation of this substrate is described in "Materials" above.

The assay is performed by mixing 150µl water, 5ml dithiothreitol (100mM, freshly prepared in water), 10µl bovine serum albumin (10mg/ml in water), 15µl NADPH (25mM, freshly prepared in 0.1M Tricine-HCl, pH 8.2), 25µl spinach ferredoxin (2mg/ml Sigma Type III in water), 3µl NADPH:ferredoxin oxidoreductase (2.5 units/ml from Sigma), and 1 µl bovine liver catalase (800,000 units/ml from Sigma); after 10 min at room temperature, this mixture is added to a 13x100 mm screw-cap test tube containing 250µl sodium 1,4-piperazinediethanesulfonate (0.1M, pH 6.0). Finally, 10µl of the sample to be assayed is added and the reaction is started by adding 30µl of the substrate, [9,10(n)-^{3}H]stearoyl-ACP (100µCi/µmol, 10µM in 0.1M sodium 1,4-piperazinediethanesulfonate, pH 5.8). After sealing with a cap, the reaction is allowed to proceed for 10 min. while shaking at 23°C. The reaction is terminated by addition of 1.2ml of 5.8% tricholoracetic acid and the resulting precipitated acyl-ACP's are removed by centrifugation. The tritium released into the aqueous supernatant by the desaturase reaction is measured by liquid scintillation spectrometry. One unit of activity is defined as the amount of enzyme required to convert 1µmol of stearoyl-ACP to oleoyl-ACP, or to release 4µg-atoms of ^{3}H per minute.

Source tissue: Developing *C. tinctorius* seeds from greenhouse grown plants are harvested between 16 and 18 days after flowering, frozen in liquid nitrogen and stored at -70°C until extracted.

Acetone Powder: Approximately 50g of frozen seeds are ground in liquid nitrogen and sieved to remove large seed coat pieces to provide a fine embryo powder. The powder is washed with acetone on a Buchner funnel until all yellow color is absent from the filtrate. The powder is then air dried and further processed as described below, or may be stored frozen for at least a year at -70°C without loss of enzyme activity.

Acetone Powder Extract: The dried acetone powder is weighed and triturated with ten times its weight of 20mM potassium phosphate, pH 6.8; the mixture is then centrifuged at 12,000 x g for 20 minutes and decanted through a layer of Miracloth (Calbiochem, La Jolla, CA).

Ion Exchange Chromatography: The acetone powder extract is then applied to a DEAE-cellulose column (Whatman DE-52) (1.5 x 12 cm) equilibrated with 20mM potassium phosphate, pH 6.8. The pass-through and a wash with one column-volume (20ml) of buffer are pooled.

Affinity Chromatography: An affinity matrix for purification of the desaturase is prepared by reacting highly purified *E. coli* ACP, with CNBr-activated Sepharose 4B (Sigma). ACP (120mg) is reduced by treatment with 1mM dithiothreitol for 30 min on ice, and then desalted on Sephadex G-10 (Pharmacia) equilibrated with 0.1M sodium bicarbonate, pH 6.0. The treated ACP (20 ml, 6 mg/ml) is then mixed with 20ml of CNBr-activated Sepharose 4B swollen in 0.1M sodium bicarbonate, pH 7.0, and the mixture is allowed to stand at 4°C for one day. The gel suspension is then centrifuged, washed once with 0.1M sodium bicarbonate, pH 7.0, and then treated with 40ml 0.1M glycine, pH 8.0, for 4 hours at room temperature to block unreacted sites. The gel is then washed for five cycles with alternating 50ml volumes of 0.5M NaCl in 0.1M sodium acetate, pH 4.0, and 0.5M NaCl in 0.1M sodium bicarbonate, pH 6.5, to remove non-covalently bound ligand. The gel is loaded into a column (1.5 x 11.2 cm) and equilibrated in 20mM potassium phosphate, pH 6.8.

The combined fractions from the DE-52 column are applied to the column, which is subsequently washed with one column volume (20ml) of the equilibration buffer, and then with 2.5 column volumes (50ml) of 300mM potassium phosphate, pH 6.8. Fractions are assayed for protein using the BCA Protein Assay Reagent (Pierce, Rockford, IL) to make sure that all extraneous protein has been eluted. Active Δ-9 desaturase is eluted from the column with 600mM potassium phosphate, pH 6.8. Active fractions are analyzed

by polyacrylamide gel electrophoresis in sodium dodecyl sulfate (SDS-PAGE) on 0.75mm thick 8 x 12 cm mini-gels according to the method of Laemmli (*Nature* (1970) *227*:680). The running gel contains 10% acrylamide in a 30/0.8 ratio of acrylamide to cross-linker bis-acrylamide. Those fractions containing a predominant band at approximately 43 kD are pooled and stored frozen at -70°C until final purification. The yield from 50g of seed tissue is is approximately 60µg of protein as measured by amino acid analysis.

Further purification as described in Example 2 or Example 3 is then applied to the fractions pooled from the ACP-Sepharose column separation.

Example 2

In this example, a protocol for the final purification of *C. tinctorius* desaturase is described. Seeds are treated in accordance with Example 1.

Reverse-Phase HPLC: Fractions from the ACP-Sepharose column are pooled and applied to a Vydac C4 reverse-phase column (0.45 x 15 cm) equilibrated in 0.1% TFA, 7% acetonitrile. After a 10 min wash with 0.1% TFA, the column is eluted with a gradient of increasing acetonitrile (7%-70% v/v) in 0.1% TFA over a period of 45 min. The flow rate is 0.5ml/min throughout. Eluting components are monitored by absorbance at 214 nm. Δ-9 desaturase elutes at about 42 min. (approximately 50% acetonitrile); the major contaminant protein remaining from ACP-affinity chromatography elutes at about 28 min. (approximately 30% acetonitrile). The substantially homogeneous desaturase, which is no longer active, is identified by SDS-PAGE, in which it exhibits a single band corresponding to a molecular weight of approximately 43 kD. The quantity of desaturase protein in the sample may be determined by amino acid analysis.

Example 3

In this example, a protocol for the final purification of *C. tinctorius* desaturase is described. Seeds are treated in accordance with Example 1.

Reduction and Alkylation: Protein is precipitated out of the pooled fraction solutions recovered from the ACP-Sepharose column with 10% (w/v) trichloroacetic acid, washed with cold (-20°C) acetone, and resuspended in 1 ml 500mM Tris-HCl, pH 8.6, containing 6M guanidine-HCl, 10mM EDTA, and 3.2 mM dithiothreitol. After 2 hours, 3.52 µmol [^{3}H]-iodoacetic acid (64µCi/µmol, New England Nuclear) is added, and the reaction is allowed to proceed at room temperature in the dark for 2 hours, at which time the reaction is terminated by addition of 1µl (15µmol) ß-mercaptoethanol. The sample is then re-precipitated with 10% (w/v) trichloroacetic acid, and the pellet again washed with cold (-20°C) acetone and resuspended in Laemmli's SDS-sample buffer (*Nature* (1970) *227*:680).

SDS-Polyacrylamide Gel Electrophoresis: The resulting sample is boiled for 5 min. and then applied to a 1.5 mm thick, 8 x 12 cm, SDS-polyacrylamide mini-gel prepared as described by Laemmli, *supra*. The running gel contains 17.5% acrylamide in a 30:0.13 ratio of acrylamide to cross-linking bis-acrylamide. Separation is achieved by electrophoresis at 15 mA, for 2 hours at 4°C.

Blotting from SDS-gels to PVDF Membrane: Proteins are recovered from the gel by electroblotting at 5 mA/cm^2 to a four-layer sandwich of polyvinylidenedifluoride (PVDF) membrane for 2 h at 4°C in a buffer containing 10mM CAPS ("3-[cyclohexylamino]-1-propane-sulfonic acid"), pH 11. The membranes must be wetted in 50% methanol, prior to exposure to the blotting buffer. After blotting, the membrane layers are stained for 1-2 min. in 0.02% Coomassie Blue in 50% methanol, and then destained in 50% methanol. The desaturase is identified as a band corresponding to a molecular weight of about 43 kD; the major contaminant runs at or near the dye front of the gel corresponding to a molecular weight less than 20 kD.

The desaturase band on the PVDF membrane may be applied directly to the Edman sequencer (Applied Biosystems Model 477A) for determination of the N-terminal sequence of the intact protein, or for more extensive sequence determination, may be eluted from the membrane in 40% acetonitrile to recover pure desaturase in solution. Acetonitrile is removed from the eluted desaturase by evaporation on a Speed-Vac (Savant; Farmingdale, NY), and the substantially homogeneous Δ-9 desaturase is resuspended in an appropriate buffer for subsequent proteolytic digestion as described in Example 4. The quantity of desaturase protein present may be determined by amino acid analysis.

Alternatively, if the sample is to be digested with trypsin or gluC protease to generate peptides for amino acid sequence analysis, proteins may be electroblotted to nitrocellulose membranes and stained with Ponceau or amido black.

Example 4

In this example, a method for the determination of the amino acid sequence of a desaturase is described.

Reduction and Alkylation: Substantially homogenous stearoyl-ACP desaturase (See, Example 2) is reduced and alkylated with [^{3}H]-iodacetic acid (See, Example 3), except that the final acetone-washed pellet is resuspended in the appropriate buffer for subsequent proteolysis. Reduction and alkylation assures complete denaturation of the protein so that complete proteolysis can occur. The sample may be alkylated with radiolabeled iodoacetamide or with 4-vinylpyridine instead of [^{3}H]-iodacetic acid in substantially the same manner. Use of iodoacetic acid affords an alkylated sample with greater solubility, which is advantageous in subsequent sample manipulation.

Proteolysis: Substantially pure alkylated samples are digested with endoproteinase lysC. The sample is resuspended in 100 µl of 25 mM Tris-HCl, pH 8.8, containing 1 mM EDTA. Endoproteinase lysC is added to the sample in a

protease/desaturase ratio of 1/50 (w/w). Digestion is allowed to proceed at room temperature for 8 hours, at which time another equal amount of protease is added. After 18 more hours, 1 µl of concentrated HCl is added to stop proteolysis, and the sample is applied directly to a Vydac C18 reverse-phase column (0.2 x 15 cm) equilibrated in 7% acetonitrile (v/v) in 0.1 mM sodium phosphate, pH 2.2. After washing for 20 min with the equilibration buffer, peptides are eluted with a gradient in acetonitrile (7-70%, v/v) over 120 min. Flow rate is 50 µl/min, throughout. Eluting components are monitored by absorbance at 214 nm, and individual peptide peaks are collected as separate fractions. The peptide fractions are further purified by application to a second Vydac C18 reverse-phase column (0.2 x 15 cm) equilibrated in 7% (v/v) acetonitrile in 0.1% (v/v) trifluoroacetic acid. Again, after a 20 min wash with equilibration buffer, the substantially pure peptides are eluted with a gradient (7-70%, v/v) of acetonitrile in 0.1% trifluoroacetic acid over 120 min. The flow rate is 50 µl/min, throughout. Eluting components are monitored by absorbance at 214 nm, and individual peptide peaks are collected as separate fractions. These substantially pure peptides are applied directly to the Edman sequencer (Applied Biosystems, Model 477A) for amino acid sequence analysis. Alternatively, peptide fraction from the first HPLC purification in phosphate buffer, or from a single chromatography step in trifluoroacetic acid buffer, may be applied directly to the sequencer, but these fractions, in many cases, are not substantially pure and yield mixed or ambiguous sequence information.

Other proteases may be used to digest desaturase, including but not limited to trypsin, gluC, and aspN. While the individual digest buffer conditions may be different, the protocols for digestion, peptide separation, purification, and sequencing are substantially the same as those outlined for the digestion with lysC. Alternatively, desaturase may be digested chemically using cyanogen bromide (Gross *Methods Enzymol* (1967) *11*:238-255 or Gross

and Witkop *J. Am. Chem. Soc.* (1961) *83*:1510), hydroxylamine (Bornstein and Balian *Methods Enzymol.* (1977) *47*:132-745), iodosobenzoic acid (Inglis *Methods Enzymol.* (1983) *91*:324-332), or mild acid (Fontana *et al.*, *Methods Enzymol.* (1983) *91*:311-317), as described in the respective references.

Fragments generated from these digestion steps of *C. tinctorius* desaturase are presented in Fig. 1 and as SEQ ID NOS: 1-11.

Example 5

In this example, the preparation of a plant embryo cDNA bank, using the methods as described in Alexander, *et al.* (*Methods in Enzymology* (1987) *154*:41-64) and the screening of the bank to obtain a desaturase cDNA clone is described.

C. tinctorius: A plant embryo cDNA library may be constructed from poly(A)+ RNA isolated from *C. tinctorius* embryos collected at 14-17 days post-anthesis. Poly(A)+ RNA is isolated from polyribosomes by a method initially described by Jackson and Larkins (*Plant Physiol.* (1976) *57*:5-10) as modified by Goldberg *et al.* (*Developmental Biol.* (1981) *83*:201-217).

The plasmid cloning vector pCGN1703, derived from the commercial cloning vector Bluescribe M13- (Stratagene Cloning Systems; San Diego, CA), is made as follows. The polylinker of Bluescribe M13- is altered by digestion with *Bam*HI, treatment with mung bean endonuclease, and blunt-end ligation to create a *Bam*HI-deleted plasmid, pCGN1700. pCGN1700 is digested with *Eco*RI and *Sst*I (adjacent restriction sites) and annealed with synthetic complementary oligonucleotides having the sequences 5' CGGATCCACTGCAGTCTAGAGGGCCCGGGA 3'(SEQ ID NO: 30) and 5' AATTTCCCGGGCCCTCTAGACTGCAGTGGATCCGAGCT 3' (SEQ ID NO: 31). These sequences are inserted to eliminate the *Eco*RI site, move the *Bam*HI site onto the opposite side of the *Sst*I (also, sometimes referred to as "*Sac*I" herein) site found in Bluescribe, and to include new restriction sites *Pst*I, *Xba*I, *Apa*I, *Sma*I. The resulting plasmid pCGN1702, is

digested with *Hin*dIII and blunt- nded with Klenow enzyme; the linear DNA is partially digested with *Pvu*II and ligated with T4 DNA ligase in dilute solution. A transformant having the *lac* promoter region deleted is selected (pCGN1703) and is used as the plasmid cloning vector.

Briefly, the cloning method for cDNA synthesis is as follows. The plasmid cloning vector is digested with *Sst*I and homopolymer T-tails are generated on the resulting 3'-overhang sticky-ends using terminal deoxynucleotidyl transferase. The tailed plasmid is separated from undigested or un-tailed plasmid by oligo(dA)-cellulose chromatography. The resultant vector serves as the primer for synthesis of cDNA first strands covalently attached to either end of the vector plasmid. The cDNA-mRNA-vector complexes are treated with terminal transferase in the presence of deoxyguanosine triphosphate, generating G-tails at the ends of the cDNA strands. The extra cDNA-mRNA complex, adjacent to the *Bam*HI site, is removed by *Bam*HI digestion, leaving a cDNA-mRNA-vector complex with a *Bam*HI sticky-end at one end and a G-tail at the other. This complex is cyclized using the annealed synthetic cyclizing linker, 5'-GATCCGCGGCCGCGAATTCGAGCTCCCCCCCCCC-3' and

3'-GCGCCGGCGCTTAAGCTCGA-5'

which has a *Bam*HI sticky-end and a C-tail end. Following ligation and repair the circular complexes are transformed into *E. coli* strain DH5α (BRL; Gaithersburg, MD) to generate the cDNA library. The *C. tinctorius* embryo cDNA bank contains between 3×10^6 and 5×10^6 clones with an average cDNA insert size of approximately 1000 base pairs.

Probe production Including PCR Reactions: Two regions of amino acid sequence (Example 4) with low codon degeneracy are chosen from opposite ends of peptide sequence "Fragment F2" (SEQ ID NO:2) for production of a probe for the plant desaturase cDNA. Two sets of mixed oligonucleotides are designed and synthesized for use as forward (SEQ ID NOS: 21-24) and reverse (SEQ ID NOS: 25-26) primers, respectively, in the polymerase chain reaction

(Saiki *et al.*, *Science* (1985) *230*:1350-1354; Oste, *Biotechniques* (1988) *6*:162-167). See, Fig. 6. All oligonucleotides are synthesized on an Applied Biosystems 380A DNA synthesizer.

Probes to *C. tinctorius* desaturase may be prepared using the peptide sequence "Fragment 2" (SEQ ID NO: 2) identified in Fig. 1. Four types of forward primers were synthesized and labeled 13-1, 13-2, 13-3, and 13-4 (SEQ ID NOS: 21-24, respectively). Two groups of reverse primers were synthesized and designated 13-5A and 13-6A (SEQ ID NOS: 25-26, respectively). The primer sequences are shown in Fig. 6. These oligonucleotide groups have a redundancy of 64 or less and contain either 20 or 17 bases of coding sequence along with flanking restriction site sequences for *Hin*dIII or *Eco*RI. Based on the intervening amino acid sequence between the primer regions on peptide "Fragment 2" (SEQ ID NO: 2) the PCR product is expected to contain 107 base pairs.

Polymerase chain reaction is performed using the cDNA library DNA as template and the possible eight combinations of the four forward and two reverse oligonucleotides as primers in a Perkin-Elmer/Cetus DNA Thermal Cycler (Norwalk, CT) thermocycle file 1 min. 94°C, 2 min. 42°C, 2 min rise from 42°-72°C for 30 cycles, followed by the step cycle file without step rises, 1 min. 94°C, 2 min. 42°C, 3 min. 72°C with increasing 15 sec extensions of the 72°C step for 10 cycles, and a final 10 min. 72°C extension.

The product of the 13-4 forward primer (SEQ ID NO: 24) and the 13-5A reverse primer (SEQ ID NO: 25) reaction was ethanol precipitated and then digested with *Hin*dIII and *Eco*RI, the resulting fragment was subcloned into pUC8 (Vieira and Messing, *Gene* (1982) *19*:259-268). Minipreparation DNA (Maniatis *et al.*, *Molecular_Cloning: A Laboratory Manual* (1982) Cold Harbor Laboratory, New York) of one clone was sequenced by Sanger dideoxy sequencing (Sanger *et al.*, *Proc. Nat. Acad. Sci. USA* (1977) *74*:5463-5467) using the M13 universal and reverse primers. Translation of the resulting DNA sequence results in a

peptide sequence that exactly matches the amino acid sequence in peptide "Fragment F2" (SEQ ID NO: 2).

An exact 50 base oligonucleotide designated DESAT-50 is synthesized to match the sequence of the PCR reaction product from the first valine residue to the last tyrosine residue.

The probe DSAT-50 5' - GTAAGTAGGTAGGGCTTCCTCTGTAATCATATCTCCAACCAAAACAACAA -3' (SEQ ID NO: 32) is used to probe the *C. tinctorius* embryo cDNA library.

Library screen

The *C. tinctorius* embryo cDNA bank is moved into the cloning vector lambda gt10 (Stratagene Cloning Systems) by digestion of total cDNA with *Eco*RI and ligation to lambda gt10 DNA digested with *Eco*RI. The titer of the resulting library was ~5x10^5/ml. The library is then plated on *E. coli* strain C600 (Huynh, *et al.*, *DNA Cloning Vol. 1* Eds. Glover D.M. IRL Press Limited: Oxford England, pp. 56, 110) at a density of 5000 plaques/150 mm NZY ("NZYM" as defined in Maniatis *et al. supra*) agar plate to provide over 45,000 plaques for screening. Duplicate lifts are taken of the plaques using NEN Colony Plaque Screen filters by laying precut filters over the plates for ~1 minute and then peeling them off. The phage DNA is immobilized by floating the filters on denaturing solution (1.5M NaCl, .05M NaOH) for 1 min., transferring the filters to neutralizing solution (1.5M NaCl, 0.5M Tris-HCI pH 8.0) for 2 min. and then to 2XSSC (1xSSC = 0.15M NaCl; 0.015M Na citrate) for 3 min., followed by air drying. The filters are hybridized with ^{32}P end-labeled DSAT-50 oligonucleotide (SEQ ID NO: 32) (BRL 5' DNA Terminus Labeling System) by the method of Devlin et al., (*DNA* (1988) 7:499-807) at 42° C overnight, and washed for 30 min. at 50°C in 2XSSC, 0.5% SDS and then twice for 20 min. each at 50°C in 0.1XSSC, 0.5% SDS. Filters are exposed to X-ray film at -70°C with a Dupont Cronex intensifying screen for 48 hours.

Clones are detected by hybridization with the DSAT-50 oligonucleotide and plaque purified. The complete nucleotide sequence (SEQ ID NO: 12) of the cDNA insert of a clone, pCGN2754, and a partial restriction map thereof are presented in Figures 2 and 7A, respectively. The cDNA insert includes 1533 bases plus a poly(A) track at the 3' end of 100-200 bases. The open reading frame for the desaturase begins at the first ATG (nucleotide 106) from the 5' end and stops at nucleotide 1294. The translated amino acid sequence is presented in Fig. 2 and SEQ ID NO: 13. The open reading frame includes a 33 amino acid transit peptide not found in the amino acid sequence of the mature protein. The N-terminus of the protein begins at the alanine immediately following the *Nco*I site (nucleotide 202) indicating the site of the transit peptide processing.

Example 6

In this example, expression of a plant desaturase in a prokaryote is described.

Desaturase expression construct in E. coli

A plasmid for expression of desaturase activity in *E. coli* is constructed as follows. The desaturase cDNA clone pCGN2754 is digested with *Hin*dIII and SalI and ligated to pCGN2016 (a chloramphenicol resistant version of Bluescript KS-) digested with *Hin*dIII and *Xho*I. The resulting plasmid is pCGN1894.

pCGN2016 is prepared by digesting pCGN565 with *Hha*I, and the fragment containing the chloramphenicol resistance gene is excised, blunted by use of mung bean nuclease, and inserted into the *Eco*RV site of Bluescript KS- (Stratagene: La Jolla, CA) to create pCGN2008. The choramphenicol resistance gene of pCGN2008 is removed by *Eco*RI/*Hin*dIII digestion. After treatment with Klenow enzyme to blunt the ends, the fragment is ligated to *Dra*I digested Bluescript KS-. A clone that has the *Dra*I fragment containing ampicillin resistance replaced with the chloramphenicol resistance is chosen and named pCGN2016.

pCGN565 is a cloning vector based on pUC12-cm (K. Buckley Ph.D. Thesis, Regulation and expression of the phi X174 lysis gene, University of California, San Diego, 1985), but contains pUC18 linkers (Yanisch-Perron, *et al.*, *Gene* (1985) *53*:103-119).

The fragment containing the mature coding region of the Δ-9 desaturase, 3'-noncoding sequences and poly(A) tails is subcloned from pCGN1894 digested with *Nco*1 and *Asp*718 into pUC120, an *E. coli* expression vector based on pUC118 (Vieira and Messing, *Methods in Enzymology* (1987) *153*:3-11) with the lac region inserted in the opposite orientation and an *Nco*I site at the ATG of the lac peptide (Vieira, J. PhD. Thesis, University of Minnesota, 1988). The *E. coli* desaturase expression plasmid is designated pCGN3201. The desaturase sequences are inserted such that they are aligned with the lac transcription and translation signals.

Expression of Desaturase in E.coli

Single colonies of *E. coli* strain 7118 (Maniatis *et al.*, *supra*) containing pUC120 or pCGN3201 are cultured in 80 mls each of ECLB broth, 300 mg/L penicillin. The cells are induced by the addition of 1mM IPTG. Cells are grown overnight (18 hrs) at 37° C.

Eighty mls of overnight cultures of *E. coli* (induced and uninduced) containing pUC120 or pCGN3201 are centrifuged at 14,800 x g for 15 min. The pelleted cells are resuspended in 3 mls 20 mM phosphate buffer, pH 6.8. Resuspended cells were broken in a french press at 16,000 psi. Broken cell mixtures are centrifuged 5000xg for 5 min. 100 µl of each supernatant is applied to a G-25 Sephadex gel filtration centrifugal column (Boehringer Mannheim Biochemicals), equilibrated in 20mM phosphate buffer pH 6.8. Columns are spun for 4 min at 5000xg. Effluent was collected and used as enzyme source in the desaturase assay.

Desaturase activity is assayed as described in Example 1. Both pUC120-containing, IPTG-induced cells and uninduced cells do not express detectable stearoyl-ACP

desaturase activity. The pCGN3201 IPTG-induced extract contains 8.22 nmol/min of desaturase activity. pCGN3201 uninduced extracts contains 6.45 nmol/min of activity. The pCGN3201 IPTG-induced extract shows 21.5% more activity than the uninduced pCGN3201 extract.

Detection of induced protein in E. coli.

Extracts of overnight cultures of *E. coli* strain 7118 (Maniatis *et al. supra*) containing pCGN3201 or pUC120 grown in ECLB containing 300 mg/L penicillin induced with 1mM IPTG are prepared as follows. 1.5 ml of overnight culture grown shaking at 37°C are pelleted in Eppendorf tubes for 10 min at 10-13,000 µg. Pellets are resuspended in 150 ul SDS sample buffer (0.05M Tris-HCl, pH6.8, 1% SDS, 5% ß-mercaptoethanol, 10% glycerol and 0.005% bromophenol blue) and boiled for 10 min. 25 µl of each sample are electrophoresed on a 10% polyacrylamide gel (Laemmli, *Nature* (1970) *227*:680) at 25 mA for 5 hours. Gels are stained in 0.05% Coomassie Brilliant Blue, 25% isopropanol and 10% acetic acid and destained in 10% acetic acid and 10% isopropanol. A band is detected at a position just below the 43,000 MW protein marker (SDS PAGE standard, Low molecular weight, BioRad, Richmond CA) in the pCGN3201 extracts that is not present in the pUC120 extracts. This is the approximate molecular weight of mature desaturase protein.

Requirement for Spinach Ferredoxin

Stearoyl-ACP desaturase can also be expressed in *E. coli* by subcloning into the *E. coli* expression vector pET8c (Studier, *et al.*, *Methods Enzymol*. (1990) *185*:60-89). The mature coding region (plus an extra Met codon) of the desaturase cDNA with accompanying 3'-sequences is inserted as an *Nco*1 - *Sma* 1 fragment into pET8c at the *Nco*1 and *Bam*H1 sites (after treatment of the *Bam*H1 site with Klenow fragment of DNA polymerase to create a blunt end) to create pCGN3208. The plasmid pCGN3208 is maintained in *E. coli* strain BL21(DE3) which contains the T7 polymerase gene

under the control of the isopropyl-b-D-thiogalactopyranoside (IPTG)-inducible *lac*UV5 promoter (Studier *et al., supra*).

E. coli cells containing pCGN3208 are grown at 37°C to an OD595 of ~0.7 in NZY broth containing 0.4% (w/v) glucose and 300 mg/liter penicillin, and then induced for 3 hours with 0.4 mM IPTG. Cells are pelleted from 1 ml of culture, dissolved in 125 µl of SDS sample buffer (10) and heated to 100°C for 10 min. Samples are analyzed by SDS polyacrylamide gel electrophoresis at 25 mA for 5 h. Gels are stained in 0.05% Coomassie Brilliant Blue, 25% (v/v) isopropanol and 10% (v/v) acetic acid. A band is detected at a position just below the 43,000 MW protein marker (SDS PAGE standard, Low Molecular Weight, BioRad, Richmond, CA) in the pCGN3208 extract that is not present in the pET8c extracts. This is the approximate molecular weight of mature desaturase protein.

For activity assays, cells are treated as described above and extracts are used as enzyme source in the stearoyl-ACP desaturase assay as described in Example 1. The extract from IPTG-induced pCGN3208 cells contains 8.61 nmol/min/mg protein of desaturase activity. The extract from pCGN3208 uninduced cells contains 1.41 nmol/min/mg protein of desaturase activity. The extract from pCGN3208 induced cells, thus has approximately 6-fold greater activity than the extract from uninduced pCGN3208 cells. Extracts from both induced and uninduced cells of pET8c do not contain detectable stearoyl-ACP desaturase activity.

Samples are also assayed as described in Example 1, but without the addition of spinach ferredoxin, to determine if the *E. coli* ferredoxin is an efficient electron donor for the desaturase reaction. Minimal activity is detected in *E. coli* extracts unless spinach ferredoxin is added exogenously.

Example 7

In this example, the preparation of an ACP expression cassette containing a plant desaturase in a binary vector suitable for plant transformation is described.

ACP Expression Cassette

An expression cassette utilizing 5'-upstream sequences and 3'-downstream sequences obtainable from *B. campestris* ACP gene can be constructed as follows.

A 1.45kb *Xho*I fragment of Bcg 4-4 (Fig. 9 and SEQ ID NO: 28) containing 5'-upstream sequences is subcloned into the cloning/sequencing vector Bluescript+ (Stratagene Cloning Systems, San Diego, CA). The resulting construct, pCGN1941, is digested with *Xho*I and ligated to a chloramphenicol resistant Bluescript M13+ vector, pCGN2015 digested with *Xho*I. pCGN2015 is prepared as described for pCGN2016 (See, Example 6) except that the *Eco*RI/*Hin*dIII "chloramphenicol" fragment isolated from pCGN2008 is ligated with the 2273 bp fragment of Bluescript KS^+ (Stratagene; LaJolla, CA) isolated after digestion with *Dra*I. This alters the antibiotic resistance of the plasmid from penicillin resistance to chloramphenicol resistance. The chloramphenicol resistant plasmid is pCGN1953.

3'-sequences of Bcg 4-4 are contained on an *Sst*I/BglII fragment cloned in the *Sst*I/*Bam*HI sites of M13 Bluescript+ vector. This plasmid is named pCGN1940. pCGN1940 is modified by *in vitro* site-directed mutagenesis (Adelman *et al.*, *DNA* (1983) *2*:183-193) using the synthetic oligonucleotide 5'-CTTAAGAAGTAACCCGGGCTGCAGTTTTAGTATTAAGAG-3' (SEQ ID NO: 33) to insert *Sma*I and *Pst*I restriction sites immediately following the stop codon of the reading frame for the ACP gene 18 nucleotides from the *Sst*I site. The 3'-noncoding sequences from this modified plasmid, pCGN1950, are moved as a *PsI-Sma*I fragment into pCGN1953 cut with *Pst*I and *Sma*I. The resulting plasmid pCGN1977 comprises the ACP expression cassette with the unique restriction sites *Eco*RV, *Eco*RI and *Pst*I available between the 1.45kb 5' and 1.5 kb of 3'-noncoding sequences (SEQ ID

NO: 28) for the cloning of genes to be expressed under regulation of these ACP gene regions.

Desaturase Expression in Plants

Desaturase cDNA sequences from pCGN2754 are inserted in the ACP expression cassette, pCGN1977, as follows. pCGN2754 is digested with *Hin*dIII (located 160 nucleotides upstream of the start codon) and *Asp*718 located in the polylinker outside the poly (A) tails. The fragment containing the coding region for desaturase was blunt-ended using DNA polymerase I and ligated to pCGN1977 digested with *Eco*RV. A clone containing the desaturase sequences in the sense orientation with respect to the ACP promoter is selected and called pCGN1895. This expression cassette may be inserted into a binary vector, for example, for *Agrobacterium*-mediated transformation, or employed in other plant transformation techniques.

Binary Vector and Agrobacterium Transformation

The fragment containing the pCGN1895 expression sequences ACP 5'/desaturase/ACP 3' is cloned into a binary vector pCGN1557 (described below) for *Agrobacterium* transformation by digestion with Asp718 and XbaI and ligation to pCGN1557 digested with Asp718 and XbaI. The resulting binary vector is called pCGN1898.

pCGN1898 is transformed into *Agrobacterium tumefaciens* strain EHA101 (Hood, *et al., J. Bacteriol.* (1986) *168*:1291-1301) as per the method of Holsters, *et al., Mol. Gen. Genet.* (1978) *163*:181-187.

RNA blot analysis of seeds (T2) from T1 plants show the presence of a mRNA species for the inserted *C. tinctorius* desaturase, but the amount of message is low compared to endogenous levels of mRNA for the *Brassica* desaturase, suggesting that the expression levels were insufficient to significantly increase the amount of desaturase enzyme above that normally present. This is consistent with the negative results from oil, desaturase activity and Western blot analyses.

Construction of pCGN1557

pCGN1557 (McBride and Summerfelt, *Plant Molecular Biology* (1990) *14(2)*:269-276) is a binary plant transformation vector containing the left and right T-DNA borders of *Agrobacterium tumefaciens* octopine Ti-plasmid pTiA6 (Currier and Nester, *supra*, the gentamycin resistance gene of pPH1JI (Hirsch and Beringer, *supra*), an *Agrobacterium rhizogenes* Ri plasmid origin of replication from pLJbB11 (Jouanin *et al.*, *supra*), a 35S promoter-kanR-tml3' region capable of conferring kanamycin resistance to transformed plants, a ColE1 origin of replication from pBR322 (Bolivar *et al.*, *supra*), and a *lacZ*' screenable marker gene from pUC18 (Yanish-Perron *et al.*, *supra*).

There are three major intermediate constructs used to generate pCGN1557:

pCGN1532 (see below) contains the pCGN1557 backbone, the pRi plasmid origin of replication, and the ColE1 origin of replication.

pCGN1546 (see below) contains the CaMV35S5'-kan^R-tml3' plant selectable marker region.

pCGN1541b (see below) contains the right and left T-DNA borders of the *A. tumefaciens* octopine Ti-plasmid and the *lacZ*' region from pUC19.

To construct pCGN1557 from the above plasmids, pCGN1546 is digested with *Xho*I, and the fragment containing the CaMV 35S5'-kan^R-tml3' region is cloned into the *Xho*I site of pCGN1541b to give the plasmid pCGN1553, which contains T-DNA/left border/CaMV 35S5'-kanR-tml3'/*lacZ*'/T-DNA left border. pCGN1553 is digested with *Bgl*II, and the fragment containing the T-DNA/left border/CaMV35S5'-kan^R-tml3'/*lacZ*'/T-DNA left border region is ligated into *Bam*HI-digested pCGN1532 to give the complete binary vector, pCGN1557.

pCGN1532

The 3.5kb *Eco*RI-*Pst*I fragment containing the gentamycin resistance gene is removed from pPh1JI (Hirsch and Beringer, *Plasmid* (1984) *12*:139-141) by *Eco*RI-*Pst*I

dig stion and cloned into *Eco*RI-*Pst*I digested pUC9 (Vieira and Messing, *Gene* (1982) *19*:259-268) to generate pCGN549. *Hin*dIII-*Pst*I digestion of pCGN549 yields a 3.1 kb fragment bearing the gentamycin resistance gene, which is made blunt ended by the Klenow fragment of DNA polymerase I and cloned into *Pvu*II digested pBR322 (Bolivar *et al., Gene* (1977) *2*:95-113) to create pBR322Gm. pBR322Gm is digested with *Dra*I and *Sph*I, treated with Klenow enzyme to create blunt ends, and the 2.8 kb fragment cloned into the Ri origin-containing plasmid pLJbB11 (Jouanin *et al., Mol. Gen. Genet.* (1985) *201*:370-374) which has been digested with *Apa*I and made blunt-ended with Klenow enzyme, creating pLHbB11Gm. The extra ColE1 origin and the kanamycin resistance gene are deleted from pLHbB11Gm by digestion with *Bam*HI followed by self closure to create pGmB11. The *Hin*dII site of pGmB11 is deleted by *Hin*dIII digestion followed by treatment with Klenow enzyme and self closure, creating pGmB11-H. The *Pst*I site of pGmB11-H is deleted by *Pst*I digestion followed by treatment with Klenow enzyme and self closure, creating pCGN1532.

<u>Construction of pCGN1546</u>

The 35S promoter-tml3' expression cassette, pCGN986, contains a cauliflower mosaic virus 35S (CaMV35) promoter and a T-DNA tml 3'-region with multiple restriction sites between them. pCGN986 is derived from another cassette, pCGN206, containing a CaMV35S promoter and a different 3' region, the CaMV region VI 3'-end. The CaMV 35S promoter is cloned as an *Alu*I fragment (bp 7144-7734) (Gardner *et. al., Nucl. Acids Res.* (1981) *9*:2871-2888) into the *Hinc*II site of M13mp7 (Messing, *et. al., Nucl. Acids Res.* (1981) *9*:309-321) to create C614. An *Eco*RI digest of C614 produced the *Eco*RI fragment from C614 containing the 35S promoter which is cloned into the *Eco*RI site of pUC8 (Vieira and Messing, *Gene* (1982) *19*:259) to produce pCGN147.

pCGN148a containing a promoter region, selectable marker (KAN with 2 ATG's) and 3' region, is prepared by

digesting pCGN528 with *Bgl*II and inserting the *Bam*HI-*Bgl*II promoter fragment from pCGN147. This fragment is cloned into the *Bgl*II site of pCGN528 so that the *Bgl*II site is proximal to the kanamycin gene of pCGN528.

The shuttle vector used for this construct pCGN528, is made as follows: pCGN525 is made by digesting a plasmid containing Tn5 which harbors a kanamycin gene (Jorgenson *et. al., Mol. Gen. Genet.* (1979) *177*:65) with *Hin*dIII-*Bam*HI and inserting the *Hin*dIII-*Bam*HI fragment containing the kanamycin gene into the *Hin*dIII-*Bam*HI sites in the tetracycline gene of pACYC184 (Chang and Cohen, *J. Bacteriol.* (1978) *134*:1141-1156). pCGN526 was made by inserting the *Bam*HI fragment 19 of pTiA6 (Thomashow *et. al., Cell* (1980) *19*:729-739), modified with *Xho*I linkers inserted into the *Sma*I site, into the *Bam*HI site of pCGN525. pCGN528 is obtained by deleting the small *Xho*I fragment from pCGN526 by digesting with *Xho*I and religating.

pCGN149a is made by cloning the *Bam*HI-kanamycin gene fragment from pMB9KanXXI into the *Bam*HI site of pCGN148a. pMB9KanXXI is a pUC4K variant (Vieira and Messing, *Gene* (1982) *19*:259-268) which has the *Xho*I site missing, but contains a functional kanamycin gene from Tn903 to allow for efficient selection in *Agrobacterium*.

pCGN149a is digested with *Hin*dIII and *Bam*HI and ligated to pUC8 digested with *Hin*dIII and *Bam*HI to produce pCGN169. This removes the Tn903 kanamycin marker. pCGN565 (see pCGN2016 description) and pCGN169 are both digested with *Hin*dIII and *Pst*I and ligated to form pCGN203, a plasmid containing the CaMV 35S promoter and part of the 5'-end of the Tn5 kanamycin gene (up to the *Pst*I site, Jorgenson *et. al.*, (1979), *supra*). A 3'-regulatory region is added to pCGN203 from pCGN204 (an *Eco*RI fragment of CaMV (bp 408-6105) containing the region VI 3' cloned into pUC18 (Yanisch-Perron, *et al.*, *Gene* (1985) *33*:103-119) by digestion with *Hin*dIII and *Pst*I and ligation. The resulting cassette, pCGN206, is the basis for the construction of pCGN986.

The pTiA6 T-DNA tml 3'-sequences are subcloned from the *Bam*19 T-DNA fragment (Thomashow *et al.*, (1980) *supra*) as a *Bam*HI-*Eco*RI fragment (nucleotides 9062 to 12,823, numbering as in Barker *et al.*, *Plant Mol. Biol.* (1982) *2*:335-350) and combined with the pACYC184 (Chang and Cohen (1978), *supra*) origin of replication as an *Eco*RI-*Hin*dIII fragment and a gentamycin resistance marker (from plasmid pLB41), obtained from D. Figurski) as a *Bam*HI-*Hin*dIII fragment to produce pCGN417.

The unique *Sma*I site of pCGN417 (nucleotide 11,207 of the *Bam*19 fragment) is changed to a *Sac*I site using linkers and the *Bam*HI-*Sac*I fragment is subcloned into pCGN565 to give pCGN971. The *Bam*HI site of pCGN971 is changed to an *Eco*RI site using linkers. The resulting *Eco*RI-*Sac*I fragment containing the tml 3' regulatory sequences is joined to pCGN206 by digestion with *Eco*RI and *Sac*I to give pCGN975. The small part of the Tn5 kanamycin resistance gene is deleted from the 3'-end of the CaMV 35S promoter by digestion with *Sal*I and *Bgl*II, blunting the ends and ligation with *Sal*I linkers. The final expression cassette pCGN986 contains the CaMV 35S promoter followed by two *Sal*I sites, an *Xba*I site, *Bam*HI, *Sma*I, *Kpn*I and the tml 3' region (nucleotides 11207-9023 of the T-DNA).

The 35S promoter-tml 3' expression cassette, pCGN986 is digested with *Hin*dIII. The ends are filled in with Klenow polymerase and *Xho*I linkers added. The resulting plasmid is called pCGN986X. The *Bam*HI-*Sac*I fragment of pBRX25 (see below) containing the nitrilase gene is inserted into *Bam*HI-*Sac*I digested pCGN986X yielding pBRX66.

Construction of pBRX25 is described in U.S. Letters Patent 4,810,648, which is hereby incorporated by reference. Briefly, the method is as follows: The nucleotide sequence of a 1212-bp *Pst*I-*Hinc*II DNA segment encoding the bromoxynil-specific nitrilase contains 65-bp of 5' untranslated nucleotides. To facilitate removal of a portion of these excess nucleotides, plasmid pBRX9 is digested with *Pst*I, and treated with nuclease *Bal*31. *Bam*HI linkers are added to the resulting ends. *Bam*HI-*Hinc*II

fragments containing a functional bromoxynil gene are cloned into the *Bam*HI-*Sma*I sites of pCGN565. The resulting plasmid, pBRX25, contains only 11 bp of 5' untranslated bacterial sequence.

pBRX66 is digested with *Pst*I and *Eco*RI, blunt ends generated by treatment with Klenow polymerase, and *Xho*I linkers added. The resulting plasmid pBRX68 now has a tml 3' region that is approximately 1.1kb. pBRX68 is digested with *Sal*I and *Sac*I, blunt ends generated by treatment with Klenow polymerase and *Eco*RI linkers added. The resulting plasmid, pCGN986XE is a 35S promoter - tml 3' expression cassette lacking the nitrilase gene.

The Tn5 kanamycin resistance gene is then inserted into pCGN986XE. The 1.0 kb *Eco*RI fragment of pCGN1536 (see pCGN1547 description) is ligated into pCGN986XE digested with *Eco*RI. A clone with the Tn5 kanamycin resistance gene in the correct orientation for transcription and translation is chosen and called pCGN1537b. The 35S promoter Kan^R-tml 3' region is then transferred to a chloramphenical resistant plasmid backbone. pCGN786, (a pUC-CAM based vector with the synthetic oligonucleotide 5' GGAATTCGTCGACAGATCTCTGCAGCTCGAGGGATCCAAGCTT 3' (SEQ ID NO: 34) containing the cloning sites *Eco*RI, *Sal*I, *Bgl*II, *Pst*I, *Xho*I, *Bam*HI, and *Hin*dIII inserted into pCGN566, pCGN566 contains the *Eco*HI-*Hin*dIII linker of pUC18 inserted into the *Eco*KI-*Hin*dIII sites of pUC13-cm (K. Buckler (1985) *supra*)) is digested with *Xho*I and the *Xho*I fragment of pCGN1537b containing the 35S promoter - Kan^R-tml 3' region is ligated in. The resulting clone is termed pCGN1546.

pCGN1541b

pCGN565RBα2X (see below) is digested with *Bgl*II and *Xho*I, and the 728bp fragment containing the T-DNA right border piece and the *lac*Z' gene is ligated with *Bgl*II-*Xho*I digested pCGN65ΔKX-S+K (see below), replacing the *Bgl*II-*Xho*I right border fragment of pCGN65ΔKX-S+K. The resulting plasmid, pCGN65α2X contains both T-DNA borders and the *lac*Z' gene. The *Cla*I fragment of pCGN65α2X is

replaced with an *Xho*I site by digesting with *Cla*I blunting the ends using the Klenow fragment, and ligating with *Xho*I linker DNA, resulting in plasmid pCGN65α2XX. pCGN65α2XX is digested with *Bgl*II and *Eco*RV, treated with the Klenow fragment of DNA polymerase I to create blunt ends, and ligated in the presence of *Bgl*II linker DNA, resulting in pCGN65α2XX'. pCGN65α2XX' is digested with *Bgl*II and ligated with *Bgl*II digested pCGN1538 (see below), resulting in pCGN1541a, which contains both plasmid backbones. pCGN1541a is digested with *Xho*I and religated. Ampicillin resistant, chlormaphenicol sensitive clones are chosen, which lack the pACYC184-derived backbone, creating pCGN1541b.

pCGN1538 is generated by digesting pBR322 with *Eco*RI and *Pvu*II, treating with Klenow to generate blunt ends, and ligating with *Bgl*II linkers. pCGN1538 is ampicillin resistant, tetracycline sensitive.

pCGN65ΔKX-S+K

pCGN501 is constructed by cloning a 1.85 kb *Eco*RI-*Xho*I fragment of pTiA6 (Currier and Nester, *J. Bact.* (1976) *126*:157-165) containing bases 13362-15208 (Barker *et al.*, *Plant Mo. Biol.* (1983) *2*:335-350) of the T-DNA (right border), into *Eco*RI-*Sal*I digested M13mp9 (Vieira and Messing, *Gene* (1982) *19*:259-268). pCGN502 is constructed by cloning a 1.6 kb *Hin*dIII-*Sma*I fragment of pTiA6, containing bases 602-2212 of the T-DNA (left border), into *Hin*dIII-*Sma*I digested M13mp9. pCGN501 and pCGN502 are both digested with *Eco*RI and *Hin*dIII and both T-DNA-containing fragments cloned together into *Hin*dIII digested pUC9 (Vieira and Messing, *Gene* (1982) *19*:259-268) to yield pCGN503, containing both T-DNA border fragments. pCGN503 is digested with *Hin*dIII and *Eco*RI and the two resulting *Hin*dIII-*Eco*RI fragments (containing the T-DNA borders) are cloned into *Eco*RI digested pHC79 (Hohn and Collins, *Gene* (1980) *11*:291-298) to generate pCGN518. The 1.6kb *Kpn*I-*Eco*RI fragment from pCGN518, containing the left T-DNA border, is cloned into *Kpn*I-*Eco*RI digested pCGN565 to

generate pCGN580. The *Bam*HII-*Bgl*II fragment of pCGN580 is cloned into the *Bam*HI site of pACYC184 (Chang and Cohen, *J. Bacteriol.* (1978) *134*:1141-1156) to create pCGN51. The 1.4 kb *Bam*HI-*Sph*I fragment of pCGN60 containing the T-DNA right border fragment, is cloned into *Bam*HI-*Sph*I digested pCGN51 to create pCGN65, which contains the right and left T-DNA borders.

pCGN65 is digested with *Kpn*I and *Xba*I, treated with Klenow enzyme to create blunt ends, and ligated in the presence of synthetic *Bgl*II linker DNA to create pCGN65ΔKX. pCGN65ΔKX is digested with *Sal*I, treated with Klenow enzyme to create blunt ends, and ligated in the presence of synthetic *Xho*I linker DNA to create pCGN65ΔKX-S+X.

<u>pCGN565RBα2X</u>

pCGN451 (see below) is digested with *Hpa*I and ligated in the presence of synthetic *Sph*I linker DNA to generate pCGN55. The *Xho*I-*Sph*I fragment of pCGN55 (bp13800-15208, including the right border, of *Agrobacterium tumefaciens* T-DNA; (Barker *et al.*, *Gene* (1977) *2*:95-113) is cloned into *Sal*I-*Sph*I digested pUC19 (Yanisch-Perron *et al.*, *Gene* (1985) *53*:103-119) to create pCGN60. The 1.4 kb *Hin*dIII-*Bam*HI fragment of pCGN60 is cloned into *Hin*dIII-*Bam*HI digested pSP64 (Promega, Inc.) to generate pCGN1039. pCGN1039 is digested with *Sma*I and *Nru*I (deleting bp14273-15208; (Barker *et al.*, *Gene* (1977) *2*:95-113) and ligated in the presence of synthetic *Bgl*II linker DNA creating pCGN1039ΔNS. The 0.47 kb *Eco*RI-*Hin*dIII fragment of pCGN1039ΔNS is cloned into *Eco*RI-*Hin*dIII digested pCGN565 to create pCGN565RB. The *Hin*dIII site of pCGN565RB is replaced with an *Xho*I site by digesting with *Hin*dIII, treating with Klenow enzyme, and ligating in the presence of synthetic *Xho*I linker DNA to create pCGN565RB-H+X.

pUC18 (Norrander *et al.*, *Gene* (1983) *26*:101-106) is digested with *Hae*II to release the *lac*Z' fragment, treated with Klenow enzyme to create blunt ends, and the *lac*Z'-containing fragment ligated into pCGN565RB-H+X, which had been digested with *Acc*I and *Sph*I and treated with Klenow

enzyme in such a orientation that the *lac*Z' promoter is proximal to the right border fragment; this construct, pCGN565RBα2x is positive for *lac*Z' expression when plated on an appropriate host and contains bp 13990-14273 of the right border fragment (Barker *et al.*, *Plant Mo. Biol.* (1983) *2*:335-350) having deleted the *Acc*I-*Sph*I fragment (bp 13800-13990).

pCGN451

pCGN451 contains an ocs5'-ocs3' cassette, including the T-DNA right border, cloned into a derivative of pUC8 (Vieira and Messing, *supra*). The modified vector is derived by digesting pUC8 with *Hinc*II and ligating in the presence of synthetic linker DNA, creating pCGN416, and then deleting the *Eco*RI site of pCGN416 by *Eco*RI digestion followed by treatment with Klenow enzyme and self-ligation to create pCGN426.

The ocs5'-ocs3' cassette is created by a series of steps from DNA derived from the octopine Ti-plasmid pTiA6 (Currier and Nester, *supra*). To generate the 5'end, which includes the T-DNA right border, an *Eco*RI fragment of pTiA6 (bp 13362-16202 (the numbering is by Barker, *et al.*, *(Plant Mol. Bio* (1983) *2*:335-350) for the closely related Ti plasmid pTi15955)) is removed from pVK232 (Knauf and Nester, *Plasmid* (1982) *8*:45) by *Eco*RI digestion and cloned into *Eco*RI digested pACYC184 (Chang and Cohen, *supra*) to generate pCGN15.

The 2.4kb *Bam*HI-*Eco*RI fragment (bp 13774-16202) of pCGN15 is cloned into *Eco*RI-*Bam*HI digested pBR322 (Bolivar, *et al.*, *supra*) to yield pCGN429. The 412 bp *Eco*RI-*Bam*HI fragment (bp 13362-13772) of pCGN15 is cloned into *Eco*RI-*Bam*HI digested pBR322 to yield pCGN407. The cut-down promoter fragment is obtained by digesting pCGN407 with *Xmn*I (bp 13512), followed by resection with *Bal*31 exonuclease, ligation of synthetic *Eco*RI linkers, and digestion with *Bam*HI. Resulting fragments of approximately 130 bp are gel purified and cloned into M13mp9 (Vieira and Messing, *supra*) and sequenced. A clone, I-4, in which the *Eco*RI linker has been inserted at bp 1362 between the

transcription initiation point and the translation initiation codon is identified by comparison with the sequence of de Greve, *et al.*, (*J. Mol. Appl. Genet.* (1982) *1*:499-512). The *Eco*RI cleavage site is at position 13639, downstream from the mRNA start site. The 141 bp *Eco*RI-*Bam*HI fragment of I-4, containing the cut-down promoter, is cloned into *Eco*RI-*Bam*HI digested pBR322 to create pCGN428. The 141 bp *Eco*RI-*Bam*HI promoter piece from pCGN428, and the 2.5 kb *Eco*RI-*Bam*HI ocs5' piece from pCGN429 are cloned together into *Eco*RI digested pUC19 (Vieira and Messing, *supra*) to generate pCGN442, reconstructing the ocs upstream region with a cut-down promoter section.

To generate the ocs3' end, the *Hin*dIII fragment of pLB41 (D. Figurski, UC San Diego) containing the gentamycin resistance gene is cloned into *Hin*dIII digested pACYC184 (Chang and Cohen, *supra*) to create pCGN413b. The 4.7 kb *Bam*HI fragment of pTiA6 (*supra*), containing the ocs3' region, is cloned into *Bam*HI digested pBR325 (F. Bolivar, *Gene* (1978) *4*:121-136) to create 33c-19. The *Sma*I site at position 11207 (Barker, *supra*) of 33c-19 is converted to an *Xho*I site using a synthetic *Xho*I linker, generating pCCG401.2. The 3.8 kb *Bam*HI-*Eco*RI fragment of pCGN401.2 is cloned into *Bam*HI-*Eco*RI digested pCGN413b to create pCGN419.

The ocs5'-ocs3' cassette is generated by cloning the 2.64 kb *Eco*RI fragment of pCGN442, containing the 5' region, into *Eco*RI digested pCGN419 to create pCNG446. The 3.1kb *Xho*I fragment of pCGN446, having the ocs5' region (bp 13639-15208) and ocs3' region (bp 11207-12823), is cloned into the *Xho*I site of pCGN426 to create pCGN451.

Example 8

In this example, the preparation of a Bce-4 expression cassette containing a plant desaturase is described.

The desaturase cDNA clone from pCGN2754 prepared as described in Example 5, is modified by *in vitro* mutagenesis to insert restriction sites immediately upstream of the ATG start codon and downstream of the TGA stop codon. A

single-stranded template DNA is prepared for the mutagenesis reaction from pCGN1894 (described in Example 6) as described by Messing, (*Methods in Enzymol.* (1983) *101*:20-79). Synthetic oligonucleotides are synthesized on an Applied Biosystems 380A DNA synthesizer. The oligonucleotides used are

5'-CCATTTTTGATCTTCCTCGAGCCCGGGCTGCAGTTCTTCTTCTTCTTG-3' (SEQ ID NO: 35) for the 5'mutagenesis and

5'-GCTCGTTTTTTTTTCTCTGCAGCCCGGGCTCGAGTCACAGCTTCACC -3' (SEQ ID NO: 36) for the 3'-mutagenesis; both add *Pst*I, *Sma*I and *Xho*I sites flanking the coding region. Both oligonucleotides are 5'-phosphorylated (BRL 5'-Terminus labelling kit) and used for mutagenesis with the pCGN1894 template by the procedure of Adelman *et al.* (*DNA* (1983) *2*:183-193). Alternatively, the desired restriction sites may be inserted by PCR, using the 3' oligo described above (SEQ ID NO: 36) and another oligo, 5' ACTGACTGCAGCCCGGGCTCGAGGAAGATCAAAAATGGCTCTTC 3' (SEQ ID NO: 37) for the 3' and 5' primers, respectively. The template in this polymerase chain reaction is DNA from pCGN1894. The *Xho*I fragment from the resulting clone can be subcloned into the Bce4 expression cassette, pCGN1870 (described below) at the unique *Xho*I site. This Bce4/desaturase expression cassette can then be inserted in a suitable binary vector, transformed into *Agrobacterium tumefaciens* strain EHA101 and used to transform plants as provided in Example 10.

<u>Bce-4 Expression Cassette</u>

pCGN1870 is a Bce-4 expression cassette containing 5' and 3' regulatory regions of the Bce-4 gene and may be derived from the Bce-4 sequence found in pCGN1857, which was deposited with the ATCC on March 9, 1990, and assigned accession number 68251, or by methods known to one skilled in the art from the sequence (SEQ ID NO: 27) provided in Fig. 8. The Bce 4 gene may be isolated as follows:

The *Cla*I fragment of pCGN1857, containing the Bce4 gene is ligated into *Cla*I digested Bluescript KS+

(Stratagene; La Jolla, CA), producing pCGN1864. Single stranded DNA is made from pCGN1864 and altered by *in vitro* mutagenesis using the oligonucleotides

BCE45P:

(5'GAGTAGTGAACTTCATGGATCCTCGAGGTCTTGAAAACCTAGA3') (SEQ ID NO: 38) and

BCE43P:

(5'CAATGTCTTGAGAGATCCCGGGATCCTTAACAACTAGGAAAAGG3') (SEQ ID NO: 39)

as described by Adelman *et al.* (*DNA* (1983) *2*:183-193). The oligonucleotide BSCP2 (5'GTAAGACACGACTTATCGCCACTG3') (SEQ ID NO: 40), complementary to a portion of Bluescript, is included in the reaction to improve the yield of double-stranded DNA molecules. The resulting plasmid, pCGN1866, contains *Xho*I and *Bam*HI sites (from BCE45P) immediately 5' to the Bce4 start codon and *Bam*HI and *Sma*I sites (from BCE43P) immediately 3' to the Bce4 stop codon. The *Cla*I fragment of pCGN1866, containing the mutagenized sequences, is inserted into the *Cla*I site of pCGN2016 (described in Example 6), producing pCGN1866C. The *Cla*I fragment of pCGN1866C is used to replace the corresponding wild-type *Cla*I fragment of PCGN1867 (described below) to produce pCGN1868. Bce4 coding sequences are removed by digestion of pCGN1868 with *Bam*HI and recircularization of the plasmid to produce pCGN1870. The Bce4 expression cassette, pCGN1870, contains 7.4 kb of 5' regulatory sequence and 1.9 kb of 3' regulatory sequence derived from the Bce4 genomic clone separated by the cloning sites, *Xho*I, *Bam*HI, and *Sma*I. Desaturase sequences in sense or anti-sense orientation may be inserted into the cassette via the cloning sites and the resulting construct may be employed in a plant transformation technique.

pCGN1867

The *Bam*HI and *Sma*I sites of pUC18 are removed by *Bam*HI-*Sma*I digestion and recircularizing of the plasmid, without repair of the ends, to produce pCGN1862 The *Pst*I fragment of pCGN1857, containing the Bce4 gene, is inserted into the *Pst*I site of pCGN1862 to produce pCGN1867.

Example 9

In this example, the preparation of a napin 1-2 expression cassette containing a plant desaturase is described.

Preparation of Desaturase Clone

The desaturase cDNA clone from pCGN2754 is prepared and modified as described in Example 8. The *Xho*I fragment from the resulting clone can be subcloned into the napin 1-2 expression cassette, pCGN1808 (described below) at the unique *Xho*I site. This napin 1-2/desaturase expression cassette can then be inserted into a suitable binary vector, transformed into *A. tumefaciens* strain EHA101 in a like manner as described in Example 7.

Alternatively, the desaturase safflower clone may be prepared such that restriction sites flank the translation start and stop sites, as described in Example 8, with the following modification. PCR was carried out according to manufacturer's instructions except for the initial annealing of the oligonucleotides to the template. The reaction mix was heated to 90°C for 5 min, cooled to 37°C over a one hour period, kept at 37°C for 20 min and then subjected to standard PCR cycles. The PCR product was digested with *Pst*I and ligated to pUC8 (Vieira and Messing (1982) *Gene 19*:2359-268) digested with *Pst*I to produce pCGN3220. The *Nco*I/*Sac*I fragment of pCGN3220 containing the pUC8 vector and the 5' and 3' sequences of the safflower desaturase cDNA was gel purified and ligated to the gel-purified cloned *Nco*I/*Sac*I fragment from pCGN1894 (see Example 6). The resulting plasmid pCGN3222 contains safflower desaturase cDNA sequences partially from the cDNA

clone and partially from the PCR. The regions obtained from the PCR were confirmed by DNA sequencing as being identical to the original cloned sequence.

Expression Cassettes

Napin 1-2 pCGN1808 Expression Cassette

An expression cassette utilizing 5' upstream sequences and 3' downstream sequences obtainable from *B. campestris* napin gene can be constructed as follows.

A 2.7 kb *Xho*I fragment of napin 1-2 (Fig. 10 and SEQ ID NO: 29) containing 5' upstream sequences is subcloned into pCGN789 (a pUC based vector the same as pUC119 with the normal polylinker replaced by the synthetic linker - 5'GGAATTCGTCGACAGATCTCTGCAGCTCGAGGGATCCAAGCTT 3', SEQ ID NO: 41, (which represented the polylinker *Eco*RI, *Sal*I, *Bgl*II, *Pst*I, *Xho*I, *Bam*HI, *Hin*dIII) and results in pCGN940. The majority of the napin coding region of pCGN940 was deleted by digestion with *Sal*I and religation to form pCGN1800. Single-stranded DNA from pCGN1800 was used in an *in vitro* mutagenesis reaction (Adelman *et al.*, *DNA* (1983) *2*:183-193) using the synthetic oligonucleotide 5' GCTTGTTCGCCATGGATATCTTCTGTATGTTC 3', SEQ ID NO: 42. This oligonucleotide inserted an *Eco*RV and an *Nco*I restriction site at the junction of the promoter region and the ATG start codon of the napin gene. An appropriate mutant was identified by hybridization to the oligonucleotide used for the mutagenesis and sequence analysis and named pCGN1801.

A 1.7 kb promoter fragment was subcloned from pCGN1801 by partial digestion with *Eco*RV and ligation to pCGN786 (a pCGN566 chloramphenicol based vector with the synthetic linker described above in place of the normal polylinker) cut with *Eco*RI and blunted by filling in with DNA Polymerase I Klenow fragment to create pCGN1802.

A 2.1 kb *Sal*I fragment of napin 1-2 (Fig. 10 and SEQ ID NO: 29) containing 3' downstream sequences is subcloned into pCGN789 (described above) and results in pCGN941. pCGN941 is digested with *Xho*I and *Hin*dIII and the resulting approximately 1.6 kb of napin 3' sequences are inserted

into *Xho*I-*Hin*dIII digested pCGN1802 to result in pCGN1803. In order to remove a 326 nucleotide *Hin*dIII fragment inserted opposite to its natural orientation, as a result of the fact that there are 2 *Hin*dIII sites in pCGN1803, the pCGN1803 is digested with *Hin*dIII and religated. Following religation, a clone is selected which now contains only 1.25 kb of the original 1.6 napin 3' sequence. This clone, pCGN1808 is the napin 1-2 expression cassette and contains 1.725 kb of napin promoter sequences and 1.265 kb of napin 3' sequence with the unique cloning sites *Sal*I, *Bgl*I, *Pst*I and *Xho*I in between.

Napin 1-2 pCGN3223 Expression Cassette

Alternatively, pCGN1808 may be modified to contain flanking restriction sites to allow movement of only the expression sequences and not the antibiotic resistance marker to binary vectors such as pCGN1557 (McBride and Summerfelt, *supra*). Synthetic oligonucleotides containing *Kpn*I, *Not*I and *Hin*dIII restriction sites are annealed and ligated at the unique *Hin*dIII site of pCGN1808, such that only one *Hin*dIII site is recovered. The resulting plasmid, pCGN3200 contains unique *Hin*dIII, *Not*I and *Kpn*I restriction sites at the 3'-end of the napin 3'-regulatory sequences as confirmed by sequence analysis.

The majority of the napin expression cassette is subcloned from pCGN3200 by digestion with *Hin*dIII and *Sac*I and ligation to *Hin*dIII and *Sac*I digested pIC19R (Marsh, *et al.* (1984) *Gene 32*:481-485) to make pCGN3212. The extreme 5'-sequences of the napin promoter region are reconstructed by PCR using pCGN3200 as a template and two primers flanking the *Sac*I site and the junction of the napin 5'-promoter and the pUC backbone of pCGN3200 from the pCGN1808 construct. The forward primer contains *Cla*I, *Hin*dIII, *Not*I, and *Kpn*I restiction sites as well as nucleotides 408-423 of the napin 5'-sequence (from the *Eco*RV site) and the reverse primer contains the complement to napin sequences 718-739 which include the unique *Sac*I site in the 5'-promoter. The PCR was performed using a Perkin Elmer/Cetus

thermocycler according to manufacturer's specifications. The PCR fragment is subclon d as a blunt-ended fragment into pUC8 (Vieira and M ssing (1982) *Gene 19*:259-268) and digested with *Hinc*II to give pCGN3217. Sequence of pCGN3217 across the napin insert verifies that no improper nucleotides were introduced by PCR. The napin 5-sequences in pCGN3217 are ligated to the remainder of the napin expression cassette by digestion with *Cla*I and *Sac*I and ligation to pCGN3212 digested with *Cla*I and *Sac*I. The resulting expression cassette pCGN3221, is digested with *Hin*dIII and the napin expression sequences are gel purified away and ligated to pIC20H (Marsh, *supra*) digested with *Hin*dIII. The final expression cassette is pCGN3223, which contains in an ampicillin resistant background, essentially identical 1.725 napin 5' and 1.265 3' regulatory sequences as found in pCGN1808. The regulatory regions are flanked with *Hin*dIII, *Not*I and *Kpn*I restriction sites and unique *Sal*I, *Bgl*II, *Pst*I, and *Xho*I cloning sites are located between the 5' and 3' noncoding regions.

Desaturase sequences in sense or anti-sense orientation may be inserted into a napin expression cassette via the cloning sites. The resulting construct may be employed for plant transformation. For example, one of ordinary skill in the art could also use known techniques of gene cloning, mutations, insertion and repair to allow cloning of a napin expression cassette into any suitable binary vector, such as pCGN1557 (described in Example 7) or other similar vectors.

Desaturase Expression

The coding region of the safflower desaturase contained in pCGN3222 is cloned into the pCGN3223 napin cassette by digestion with *Xho*I and ligation to pCGN3223 digested with *Xho*I and *Sal*I. The resulting plasmid, pCGN3229 is digested with *Asp*718 and inserted in the binary vector pCGN1578 (McBride and Summerfelt (1990) *Plant Mol. Biol. 14*:269-276) at the unique *Asp*718 site. The resulting binary vector is pCGN3231 and contains the safflower

desaturase coding sequences flanked by the napin 5' and 3' regulatory s quences as well as the plant selectable marker construct, 35s/NPTII/tml.

The resulting binary vector, pCGN3231, is transformed into *Agrobacterium* and utilized for plant transformation as described in Example 10. For Northern analysis, total RNA is isolated from day 21 and day 28 post-anthesis developing seed from plants transformed with pCGN3231. Five samples were analyzed at day 21 and two at day 28 post-anthesis. RNA was isolated by the method of Hughes and Galau (*Plant Mol. Biol. Reporter* (1988) *6*: 253-257). Northern blot analysis was performed using a labeled 0.8 kb *Bgl*II fragment of pCGN1894 as a probe. Prehybridization and hybridization was at 42°C in 50% formamide, 10X Denhardt's solution, 5X SSC, 0.1% SDS, 5mM EDTA and 100ug/ml denatured salmon sperm DNA. Filters were washed at 55°C in 0.1 X SSC, 0.1% SDS. Under these conditions, the probe does not hybridize to the endogenous *Brassica* desaturase gene sequences. mRNA complementary to the safflower desaturase was detected in all the transgenic samples examined. More mRNA was present at day 28 than at day 21 post-anthesis and the highest level of RNA was seen in transgenic 3231-8. The total safflower desaturase mRNA level was estimated to be ~0.01% of the message at day 28 post-anthesis.

Western analysis (see below) gives a preliminary indication of increased protein in one transformant, 3231-8. However, the Western analysis is complicated by two factors: 1. The presence of cross-reacting material at the same molecular weight as expected for the safflower desaturase. We believe this material is the endogenous *Brassica* desaturase. 2. The analysis of levels of protein expressed is also complicated by the normal developmental increase in the expression of desaturase protein during this time period. If the seeds examined are not at the precise developmental stage as the control seeds, quantitative differences in the amount of material seen may be simply due to the normal increase in the *Brassica*

desaturase over this time period and not due to the expression of the safflower desaturase.

Western Analysis

Soluble protein is extracted from developing seeds of *Brassica* by homogenization with one volume (1ml/gram fresh weight) of buffer containing 20mM potassium phosphate, pH 6.8. The homogenate is clarified by centrifugation at 12,000 x *g* for 10 minutes. A second centrifugation is performed if necessary to provide a non-particulate supernatant.

Protein concentration of the extract is measured by the micromethod of Bradford (*Anal. Biochem.* (1976) *72*:248-254). Proteins (20-60μg) are separated by denaturing electrophoresis by the method of Laemmli (*supra*), and are transferred to nitrocellulose membrane by the method of Towbin *et al.* (*Proc. Nat. Acad Sci.* (1979) *76*:4350-4354).

The nitrocellulose membrane is blocked by incubation at room temperature for 15 minutes or at 4°C overnight in Tris-buffered saline with Tween 20 (Polyoxyethylenesorbitan monolaurate) and "TTBS-milk", (TTBS = 20mM Tris-HCl, 500mM NaCl, 0.1% Tween 20 (v/v), pH 7.5; "TTBS-milk" = TTBS and 3% skim milk powder). The volume of liquid in all incubations with the nitrocellulose membrane is sufficient to cover the membrane completely. The membrane is then incubated for an additional 5 minutes in TTBS.

The nitrocellulose membrane is incubated for at least one hour with shaking at room temperature with rabbit anti-stearoyl-ACP desaturase antiserum that was diluted 5,000- or 10,000-fold in "TTBS-milk". The rabbit anti-desaturase antiserum was commercially prepared from desaturase protein (purified as described in Example 1) by Berkeley Antibody Co. (Richmond, CA). The membrane is washed twice by shaking with TTBS for 5 minutes and then with deionized H_2O for 30 seconds.

The nitrocellulose membrane is incubated for at least 45 minutes at room temperature in a solution of "TTBS-milk" in which anti-rabbit IgG-alkaline phosphatase conjugate

(Promega, Madison, WI) is diluted 7,500-fold. Th membrane is washed twice in TTBS followed by deionized H_2O, as described above.

The nitrocellulose membrane is equilibrated in buffer containing 100mM Tris-HCl, 100mM NaCl, 50mM $MgCl_2$, pH 9.5, by shaking for 5 minutes. The color reaction is initiated by placing the nitrocellulose membrane into 50ml of the same buffer to which has been added 15mg *p*-nitroblue tetrazolium chloride and 7.5mg 5-bromo- 4 chloro- 3-indolyl phosphate toluidine salt (BioRad Labs; Richmond, CA). The color reaction is stopped by rinsing the nitrocellulose membrane with deionized H_2O and drying it between filter papers.

Oil analysis of developing seeds indicated no significant change in oil composition of the transformed plants with respect to the control plants. This result is consistant with the low levels of safflower mRNA observed in transgenic plants as compared to levels of endogenous *Brassica* desaturase (Example 12).

Example 10

In this example, an *Agrobacterium*-mediated plant transformation is described. *Brassica napus* is exemplified. The method is also useful for transformation of other *Brassica* species including *Brassica campestris*.

Plant Material and Transformation

Seeds of *Brassica napus* cv. Delta are soaked in 95% ethanol for 2 min, surface sterilized in a 1.0% solution of sodium hypochlorite containing a drop of Tween 20 for 45 min., and rinsed three times in sterile, distilled water. Seeds are then plated in Magenta boxes with 1/10th concentration of Murashige minimal organics medium (Gibco) supplemented with pyrodoxine (50 µg/l), nicotinic acid (50 µg/l), glycine (200 µg/l), and 0.6% Phytagar (Gibco) pH 5.8. Seeds are germinated in a culture room at 22°C in a 16 h photoperiod with cool fluorescent and red light of

intensity approximately 65 µEinsteins per square meter per second ($\mu Em^{-2}S^{-1}$).

Hypocotyls are excised from 7 day old seedlings, cut into pieces approximately 4 mm in length, and plated on feeder plates (Horsch et al. 1985). Feeder plates are prepared one day before use by plating 1.0 ml of a tobacco suspension culture onto a petri plate (100x25 mm) containing about 30 ml MS salt base (Carolina Biological) 100 mg/l inositol, 1.3 mg/l thiamine-HCl, 200 mg KH_2PO_4 with 3% sucrose, 2,4-D (1.0 mg/l), 0.6% Phytagar, and pH adjusted to 5.8 prior to autoclaving (MS0/1/0 medium). A sterile filter paper disc (Whatman 3 mm) is placed on top of the feeder layer prior to use. Tobacco suspension cultures are subcultured weekly by transfer of 10 ml of culture into 100 ml fresh MS medium as described for the feeder plates with 2,4-D (0.2 mg/l), Kinetin (0.1 mg/l). All hypocotyl explants are preincubated on feeder plates for 24 h. at 22°C in continuous light of intensity 30 $\mu Em^{-2}S^{-1}$ to 65 $\mu EM^{-2}S^{-1}$.

Single colonies of *A. tumefaciens* strain EHA101 containing a binary plasmid are transferred to 5 ml MG/L broth and grown overnight at 30°C. Per liter, MG/L broth contains 5g mannitol, 1 g L-glutamic acid or 1.15 g sodium glutamate, 0.25 g kH_2PO_4, 0.10 g NaCL, 0.10 g $MGSO_4 \cdot 7H_2O$, 1 mg biotin, 5 g tryptone, and 2.5 g yeast extract, and the broth is adjusted to pH 7.0. Hypocotyl explants are immersed in 7-12 ml MG/L broth with bacteria diluted to $1x10^8$ bacteria/ml and after 10-20 min. are placed onto feeder plates. After 48 h of co-incubation with *Agrobacterium*, the hypocotyl explants are transferred to B5 0/1/0 callus induction medium which contains filter sterilized carbenicillin (500 mg/l, added after autoclaving) and kanamycin sulfate (Boehringer Mannheim) at concentrations of 25 mg/l.

After 3-7 days in culture at 65 $\mu Em^{-2}S^{-1}$ to 75 $\mu Em^{-2}S^{-1}$ continuous light, callus tissue is visible on the cut surface and the hypocotyl explants are transferred to shoot induction medium, B5BZ (B5 salts and vitamins supplemented

with 3 mg/l benzylaminopurine, 1 mg/l zeatin, 1% sucrose, 0.6% Phytagar and pH adjusted to 5.8). This medium also contains carbenicillin (500 mg/l) and kanamycin sulfate (25 mg/l). Hypocotyl explants are subcultured onto fresh shoot induction medium every two weeks.

Shoots regenerate from the hypocotyl calli after one to three months. Green shoots at least 1 cm tall are excised from the calli and placed on medium containing B5 salts and vitamins, 1% sucrose, carbenicillin (300 mg/l), kanamycin sulfate (50 mg/l) and 0.6% Phytagar) and placed in a culture room with conditions as described for seed germination. After 2-4 weeks shoots which remain green are cut at the base and transferred to Magenta boxes containing root induction medium (B5 salts and vitamins, 1% sucrose, 2 mg/l indolebutyric acid, 50 mg/l kanamycin sulfate and 0.6% Phytagar). Green rooted shoots are tested for NPT II activity.

Example 11

In this example, a DNA-bombardment plant transformation is described. Peanut transformation is exemplified.

DNA sequences of interest may be introduced as expression cassettes, comprising at least a promoter region, a gene of interest, and a termination region, into a plant genome via particle bombardment as described in European Patent Application 332 855 and in co-pending application USSN 07/225,332, filed July 27, 1988.

Briefly, tungsten or gold particles of a size ranging from 0.5µM-3µM are coated with DNA of an expression cassette. This DNA may be in the form of an aqueous mixture or a dry DNA/particle precipitate.

Tissue used as the target for bombardment may be from cotyledonary explants, shoot meristems, immature leaflets, or anthers.

The bombardment of the tissue with the DNA-coated particles is carried out using a Biolistics™ particle gun (Dupont; Wilmington, DE). The particles are placed in the barrel at variable distances ranging from 1cm-14cm from the

barrel mouth. The tissue to be bombarded is placed beneath the stopping plate; testing is performed on the tissue at distances up to 20 cm. At the moment of discharge, the tissue is protected by a nylon net or a combination of nylon nets with mesh ranging from 10μM to 300μM.

Following bombardment, plants may be regenerated following the method of Atreya, *et al.*, (*Plant Science Letters* (1984) *34*:379-383). Briefly, embryo axis tissue or cotyledon segments are placed on MS medium (Murashige and Skoog, *Physio. Plant.* (1962) *15*:473) (MS plus 2.0 mg.1 6-benzyladenine (BA) for the cotyledon segments) and incubated in the dark for 1 week at 25 ± 2°C and are subsequently transferred to continuous cool white fluorescent light (6.8 W/m^2). On the 10th day of culture, the plantlets are transferred to pots containing sterile soil, are kept in the shade for 3-5 days are and finally moved to greenhouse.

The putative transgenic shoots are rooted. Integration of exogenous DNA into the plant genome may be confirmed by various methods known to those skilled in the art.

Example 12

This example describes methods to obtain desaturase cDNA clones from other plant species using the DNA from the *C. tinctorius* Δ-9 desaturase clone as the probe.

Isolation of RNA for Northern Analysis

Poly(A)+ RNA is isolated from *C. tinctorius* embryos collected at 14-17 days post-anthesis and *Simmondsia chinensis* embryos as described in Example 5.

Total RNA is isolated from days 17-18 days post-anthesis *Brassica campestris* embryos by an RNA minipreparation technique (Scherer and Knauf, *Plant Mol. Biol.* (1987) *9*:127-134). Total RNA is isolated from *R. communis* immature endosperm of about 14-21 days post-anthesis by a method described by Halling, *et al.* (*Nucl. Acids Res.* (1985) *13*:8019-8033). Total RNA is isolated

from 10 g each of young leaves from *B. campestris*, *B. napus*, and *C. tinctorius*, by extraction of each sample in 5 ml/g tissue of 4 M guanidine thiocyanate buffer as described by Colbert *et al.* (*Proc. Nat. Acac. Sci.* (1983) *80*:2248-2252). Total RNA is also isolated from immature embryos of *Cuphea hookeriana* by extraction as above in 10 ml/g tissue.

Total RNA is isolated from immature embryos of California bay (*Umbellularia californica*) by an adaptation of the method of Lagrimini *et al.* (*Proc. Nat. Acad. Sci.* (1987) *84*:7542-7546). Following homogenization in grinding buffer (2.5 ml/g tissue) as described, RNA is precipitated from the aqueous phase by addition of 1/10 volume 3 M sodium acetate and 2 volumes ethanol, followed by freezing at -80°C for 30 minutes and centrifugation at 13,000 x g for 20 minutes. The pellets are washed with 80% ethanol and centrifugation is repeated as above. The pellets are resuspended in water, two volumes of 4 M LiCl are added, and the samples are placed at -20°C overnight. Samples are centrifuged as above and the pellets washed with 80% ethanol. Ethanol precipitation is repeated as above.

Total RNA is further purified from *B. campestris*, *B. napus*, and *C. tinctorius* leaves, and from *C. tinctorius*, *B. campestris*, California bay, and jojoba, and from *R. communis* immature endosperm, by removing polysaccharides on a 0.25 g Sigma Cell 50 cellulose column. The RNA is loaded onto the column in 1 ml of loading buffer (20 mM Tris-HCl pH 7.5, 0.5M NaCl, 1mM EDTA, 0.1% SDS), eluted with loading buffer, and collected in 500 µl fractions. Ethanol is added to the samples to precipitate the RNA. The samples are centrifuged, and the pellets resuspended in sterile distilled water, pooled, and again precipitated in ethanol. The sample is centrifuged, and the resulting RNA is subjected to oligo(dT)-cellulose chromatography to enrich for poly(A)+ RNA as described by Maniatis *et al.* (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York (1982)). Poly(A)+ RNA is also

purified from total *Cuphea hookeriana* RNA by oligo(dT)-cellulose chromatography.

Northern Analysis Using C. tinctorius Desaturase Clone: 2.5 µg of poly(A)+ RNA from each of the above described poly(A)+ samples from immature embryos of jojoba, *Cuphea hookeriana*, California bay, *Brassica campestris*, and *C. tinctorius*, from immature endosperm of *R. communis*, and from leaves of *C. tinctorius*, *B. campestris*, and *B. napus* are electrophoresed on formaldehyde/agarose gels (Fourney *et al.*, *Focus* (1988) *10*:5-7) and transferred to a Hybond-C extra (Amersham, Arlington Heights, IL) filter according to manufacturer's specifications. The filter is prehybridized for four hours and hybridized overnight at 42°C in a roller bottle containing 10 ml of hybridization buffer (1 M NaCl, 1% SDS, 50% formamide, 0.1 mg/ml denatured salmon sperm DNA) in a Hybridization Incubator, model 1040-00-1 (Robbins Scientific Corporation, Sunnyvale, CA). The probe used in the hybridization is a gel-isolated *Bgl*II fragment of the Δ-9 desaturase clone that is labeled with ^{32}P-dCTP using a BRL (Gaithersburg, MD) nick-translation kit, following manufacturer's instructions. The blot is washed three times for 20 minutes each in 2X SSC, 0.5% SDS at 55°C. The blot is exposed at -80°C, with a Dupont Cronex intensifying screen, to X-ray film for four days.

The autoradiograph shows that the *C. tinctorius* desaturase gene is expressed in both immature embryos and leaves of *C. tinctorius*, although the level of expression is considerably higher in embryos than in leaves. The autoradiograph also shows hybridization of the *C. tinctorius* desaturase clone to mRNA bands of a similar size in immature embryos from jojoba and California bay, and immature endosperm from *R. communis*. Hybridization is also detectable in RNA from *B. campestris* embryos upon longer exposure of the filter to X-ray film.

R. communis cDNA Library Construction: A plant seed cDNA library may be constructed from poly(A)+ RNA isolated from *R. communis* immature endosperm as described above. The plasmid cloning vector pCGN1703, and cloning method are

as described in Example 5. The *R. communis* endosperm cDNA bank contains approximately 2x10^6 clones with an average cDNA insert size of approximately 1000 base pairs.

The *R. communis* immature endosperm cDNA bank is moved into the cloning vector lambda gt22 (Stratagene Cloning Systems) by digestion of total cDNA with *Not*I and ligation to lambda gt22 DNA digested with *Not*I. The resulting phage are packaged using a commercially available kit and titered using *E. coli* strain LE392 (Stratagene Cloning Systems, La Jolla, CA). The titer of the resulting library was approximately 1.5 x 10^7 pfu/ml.

R. communis cDNA Library Screen: The library is plated on *E. coli* strain LE392 at a density of approximately 25,000 pfu/150mm NZY plate to provide approximately 50,000 plaques for screening. Phage are lifted in duplicate on to NEN (Boston, MA.) Colony/Plaque Screen filters as described in Example 5. Following prehybridization at 42°C in 25 ml of hybridization buffer (1 M NaCl, 1% SDS, 50% formamide, 0.1 mg/ml denatured salmon sperm DNA) filters are hybridized overnight with a gel-purified 520 base pair *Bgl*II fragment of the *C. tinctorius* desaturase clone (Figure 7A) that is radiolabeled with ^{32}P-dCTP using a BRL (Gaithersburg, MD) Nick Translation System. Filters are washed three times for 20 minutes each in 2X SSC, 0.5% SDS at 55°C in a shaking water bath. Filters are exposed to X-ray film overnight at -80°C with a Dupont Cronex intensifying screen.

Clones are detected by hybridization on duplicate filters with the *C. tinctorius* desaturase cDNA fragment and plaque purified. During plaque purification, it was observed that larger plaques were obtained when *E. coli* strain Y1090 (Young, R.A. and Davis, R.W., *Proc. Natl. Acad. Sci. USA* (1983) *80*:1194) was used as the host strain. This strain was thus used in subsequent plaque purification steps. Phage DNA is prepared from the purified clones as described by Grossberger (*NAR* (1987) *15*:6737) with the following modification. The proteinase K

treatment is replaced by the addition of 10% SDS and a 10 minute incubation at room temperature. Recovered phage DNA is digested with *Eco*RI, religated at low concentration, and transformed into *E. coli* DH5α(BRL; Gaithersburg, MD) cells to recover plasmids containing cDNA inserts in pCGN1703. Minipreparation DNA (Maniatis *et al., supra*) is prepared from the clones and DNA sequence is determined as described above. Partial nucleotide sequence of the cDNA insert of a *R. communis* desaturase clone pCGN3230 is presented in Figure 3A and SEQ ID NO: 14. The complete nucleotide sequence of this clone is presented in Fig. 3B and as SEQ ID NO: 15.

Northern Analysis Using R. communis Desaturase Clone: Total RNA for Northern analysis is isolated from tobacco leaves by the method of Ursin *et al.*(*Plant Cell* (1989) *1*:727-736), petunia and tomato leaves by the method of Ecker and Davis (*Proc.Nat.Acad.Sci.* (1987) *84*:5202-5206), and corn leaves by the method of Turpen and Griffith (*Biotechniques* (1986) *4*:11-15). Total RNA samples from tobacco, corn, and tomato leaves are enriched for poly(A)+ RNA by oligo(dT)-cellulose chromatography as described by Maniatis *et al.* (*supra*).

Poly(A)+ RNA samples from tomato leaves (4 μg) and corn and tobacco leaves (1 μg each), and total RNA from petunia leaves (25 μg) are electrophoresed on a formaldehyde/agarose gel as described by Shewmaker *et al.* (*Virology* (1985) *140*:281-288). Also electrophoresed on this gel are poly(A)+ RNA samples isolated from *B. campestris* day 17-19 embryos and *B. campestris* leaves (2 μg each), immature embryos from *C. tinctorius*, bay, and jojoba (1 μg each), and *R. communis* endosperm (1 μg). The isolation of these poly(A)+ RNA samples is described above for the Northern analysis using *C. tinctorius* desaturase cDNA as probe. The RNA is transferred to a nitrocellulose filter as described by Shewmaker *et al.* (*supra*) and prehybridized and hybridized at 42°C in 50% formamide, 10X Denhardt's solution (described in Maniatis *et al.* (supra)), 5X SSC, 0.1% SDS, 5 mM EDTA, 100 ug/ml denatured salmon

sperm DNA, and 10% dextran sulfate (in hybridization buffer only). The probe for hybridization is the ^{32}P-labeled (BRL Nick Translation Kit) 1.7 kb SalI insert of pCGN3230 that has been gel-purified from minipreparation DNA. The filter is washed following hybridization for 30 minutes in 2X SSC, 0.1% SDS at 42°C and at 50°C twice for 15 minutes each. The filter is exposed to X-ray film overnight at -80°C with a Dupont Cronex intensifying screen.

The autoradiograph shows hybridization of the *R. communis* desaturase clone to mRNA bands of a similar size in immature embryos from *B. campestris*, California bay, and *C. tinctorius*, and also in corn leaves and *R. communis* endosperm.

B. campestris Embryo cDNA Library Construction: Total RNA is isolated from 5 g of *B. campestris* cv. R500 embryos obtained from seeds harvested at days 17-19 post-anthesis. RNA is extracted in 25 mls of 4 M guanidine thiocyanate buffer as described by Colbert *et al.*(*PNAS* (1983) *80*:2248-2252). Polysaccharides are removed from the RNA sample by resuspending the pellet in 6 ml of 1X TE (10 mM Tris/1 mM EDTA pH 8), adding potassium acetate to a concentration of 0.05M, and adding one half volume of ethanol. The sample is placed on ice for 60 minutes and centrifuged for 10 minutes at 3000 x g. RNA is precipitated from the supernatant by adding sodium acetate to a concentration of 0.3 M followed by the addition of two volumes of ethanol. RNA is recovered from the sample by centrifugation at 12,000 x g for 10 minutes and yield calculated by UV spectrophotometry. Two mg of the total RNA is further purified by removing polysaccharides on a 0.25 g Sigma Cell 50 cellulose column, as described above, and is also enriched for poly(A)+ RNA by oligo(dT)-cellulose chromatography as described above.

A *B. campestris* day 17-19 post anthesis embryo cDNA library is constructed in plasmid vector pCGN1703 as described in Example 5, using 5 ug of the above described poly(A)+ RNA. The library, which consists of approximately 1.5 x 10^5 transformants, is amplified by plating and scraping colonies, and is stored as frozen *E. coli* cells in

10% DMSO at -80° C. DNA is isolated from a portion of the amplified library by scaling up the alkaline lysis technique of Birnboim and Doly (*Nucleic Acids Res.* (1979) 7:1513), and purified by CsCl centrifugation. Library DNA is digested with *Eco*RI and is cloned into *Eco*RI-digested bacteriophage lambda gt10 (Stratagene; La Jolla, CA) DNA. The DNA is packaged using Gigapack II Gold *in vitro* packaging extracts (Stratagene; La Jolla, CA) according to manufacturer's specifications. The titer of the phage stock, determined by dilution plating of phage in *E. coli* C600 hfl- cells (Huynh, *et al., DNA Cloning. Volume 1.* Eds. Gover, D.M. (1985) IRL Press Limited: Oxford, England, pp. 56,110), is 6 x 10^6 pfu per ml.

B. campestris cDNA Library Screen: The library is plated on *E. coli* strain C600 hfl- at a density of approximately 30,000 pfu/150mm NZY plate to provide approximately 120,000 plaques for screening. Phage are lifted in duplicate on to NEN (Boston, MA.) Colony/Plaque Screen filters as described in Example 5. Filters are prehybridized and hybridized with the ^{32}P-labeled fragment of pCGN3230 as described above for the Northern hybridization. Filters are washed for 30 minutes in 2X SSC, 0.1% SDS at 50°C and at 55°C twice for 15 minutes each. Filters are exposed to X-ray film overnight at -80°C with a Dupont Cronex intensifying screen.

Clones are detected by hybridization on duplicate filters to the *R. communis* desaturase cDNA fragment and plaque purified. During plaque purification, the probe used was a gel-purified 1.4 kb SstI fragment of pCGN3230 which lacks the poly(A)+ tail. As described above, phage DNA is isolated from purified lambda clones, digested with EcoRI, ligated, and transformed to *E. coli* DH5α cells. Minipreparation DNA is prepared and partial DNA sequence determined as described above. Partial DNA sequences of two clones, pCGN3235 and pCGN3236, are presented in Figure 4A (SEQ ID NO: 17) and 4B (SEQ ID NO: 18), respectively. Initial DNA sequence analysis of the 3' regions of these clones indicates that pCGN3236 and pCGN3235 are cDNA

clones from the same gene. pCGN3236 is a shorter clon than pCGN3235, which appears to contain the entire coding region of the *B. campestris* desaturase gene. The complete nucleotide sequence of pCGN3235 is presented in Figure 4C and SEQ ID NO: 19.

Desaturase Gene Analysis: Southern and Northern analyses of Brassica species are conducted to determine the number of genes which encode the Brassica desaturase clone, pCGN3235 in *B. campestris*, *B. oleracea*, and *B. napus*, and the timing of expression of the gene in *B. campestris* developing seeds. DNA is isolated from leaves of each of the above-named *Brassica* species by the method of Bernatzky and Tanksley (*Theor. Appl. Genet.* (1986) *72*:314-321). DNA from each of the species is digested with restriction endonucleases EcoRI and XbaI (10 ug/digest), electrophoresed in a 0.7% agarose gel, and transferred to a nitrocellulose filter (Maniatis *et al.*, *supra*). The filter is prehybridized and hybridized at 42°C (as described above for Northern analysis using *R. communis* desaturase clone) with a ^{32}P-labeled (nick translation) gel-isolated HindIII/PvuII fragment of pCGN3235 (Fig. 7C). The filter is washed following overnight hybridization, for 30 minutes at 55°C in 1X SSC, 0.1% SDS, followed by two 15 minute washed at 55°C in 0.1X SSC, 0.1% SDS.

The autoradiograph indicates that the *Brassica* desaturase is encoded by a small gene family consisting of about two genes in *B. campestris* and *B. oleracea*, and about four genes in *B. napus*.

The timing of expression of the desaturase gene during seed development is determined by Northern analysis. RNA is isolated from immature seeds of *B. campestris* cv. R500 collected at 11, 13, 15, 17, 19, 21, 25, 30, 35, and 40 days post-anthesis. Total RNA is isolated as described by Scherer and Knauf (*Plant Mol. Biol.* (1987) *9*:127-134). Twenty five micrograms of RNA from each time point are electrophoresed through a formaldehyde-containing 1.5% agarose gel as described by Shewmaker, *et al.* (*supra*) and blotted to nitrocellulose (Thomas, *Proc. Nat. Acad. Sci.*

(1980) *77*:5201-5205). The blot is pre-hybridized and hybridized at 42°C with the ^{32}P-labeled HindIII/PvuII fragment of pCGN3235 as described above. The filter is washed following overnight hybridization, for 30 minutes at 55°C in 1X SSC, 0.1% SDS, followed by two 15 minute washed at 55°C in 0.1X SSC, 0.1% SDS.

The autoradiograph indicates that the desaturase gene is expressed in *B. campestris* developing seeds beginning at about day 19 and through about day 30, with maximal expression at day 25. By a similar Northern analysis, the level of desaturase mRNA in developing *Brassica napus* seeds (day 21) was estimated to be approximately 1% of the total mRNA.

Isolation of Other Desaturase Gene Sequences: cDNA libraries may be constructed as described above and genomic libraries can be constructed from DNA from various sources using commercially available vectors and published DNA isolation, fractionation, and cloning procedures. For example, a *B. campestris* genomic library can be constructed using DNA isolated according to Scofield and Crouch (J.Biol.Chem. (1987) 262:12202-12208) that is digested with BamHI and fractionated on sucrose gradients (Maniatis *et al., supra*), and cloned into the lambda phage vector LambdaGem-11 (Promega; Madison,WI) using cloning procedures of Maniatis *et al.* (*supra*).

cDNA and genomic libraries can be screened for desaturase cDNA and genomic clones, respectively, using published hybridization techniques. Screening techniques are described above for screening libraries with DNA fragments. Libraries may also be screened with synthetic oligonucleotides, for example using methods described by Berent *et al.* (*BioTechniques* (1985) *3*:208-220). Probes for the library screening can be prepared by PCR, or from the sequences of the desaturase clones provided herein. Oligonucleotides prepared from the desaturase sequences may be used, as well as longer DNA fragments, up to the entire desaturase clone.

For example, jojoba polyadenylated RNA is used to construct a cDNA library in the cloning vector λZAPII/*Eco*RI (Stratagene, San Diego, CA). RNA is isolated from jojoba embryos collected at 80-90 days post-anthesis by isolating polyribosomes using a method initially described by Jackson and Larkins (*Plant Physiol.* (1976) *57*:5-10) and modified by Goldberg *et al.* (*Developmental Biol.* (1981) *83*:201-217). Polysaccharide contaminants in the polyribosomal RNA preparation are removed by running the RNA over a cellulose column (Sigma-cell 50) in high salt buffer (0.5M NaCl, 20mM Tris pH 7.5, 1mM EDTA, 0.1% SDS). The contaminant binds to the column and the RNA is collected in the eluant. The eluant fractions are pooled and the RNA is ethanol precipitated. The precipitated total RNA is then resuspended in a smaller volume and applied to an oligo d(T) cellulose column to isolate the polyadenylated RNA.

The library is constructed using protocols, DNA and bacterial strains as supplied by the manufacturer. Clones are packaged using Gigapack Gold packaging extracts (Stratagene), also according to manufacturer's recommendations. The cDNA library constructed in this manner contins approximately 1 x 10^6 clones with an average cDNA insert size of approximately 400 base pairs.

The jojoba library is plated on *E. coli* XL1-Blue (Stratagene) at a density of approximately 5000pfu/150mm plate to provide approximately 60,000 plaques for screening. Phage are lifted onto duplicate nylon membrane filters as described previously. Filters are prehybridized at 42°C in a hybridization buffer containing 40% formamide, 10X Denhardt's solution, 5X SSC, 0.1% SDS, 50mM EDTA, and 100μg/ml denatured salmon sperm DNA. Hybridization is at 42°C in the same buffer with added nick translated (BRL Nick Translation System) 520 bp *Bgl*II fragment of the *C. tinctorius* desaturase clone described previously. Filters are washed at 50°C in 2X SSC and exposed to X-ray film overnight.

Desaturase clones are detected by hybridization on duplicate filters with the *C. tinctorius* cDNA fragment and

plaque-purified. Positive clones are recovered as plasmids in *E. coli* following manufacturer's directions and materials for *in vivo* excision. Partial, preliminary DNA sequence of a clone, 3-1, is determined and the corresponding amino acid sequence is translated in three frames. In this manner, homology to the *C. tinctorius* desaturase cDNA clone is detected in one reading frame. The preliminary DNA sequence of this jojoba desaturase cDNA fragment is shown in Figure 5 (SEQ ID NO: 43). Also shown is the corresponding translated amino acid sequence in the reading frame having *C. tinctorius* desaturase homology. The jojoba cDNA fragment is approximately 75% homologous at the DNA level and approximately 79% homologous at the amino acid level compared to sequence of the *C. tinctorius* desaturase in this region.

Example 13

Antisense constructs are described which allow for transcription of a reverse copy of the *B. campestris* desaturase cDNA clone in the 5' to 3' orientation of transcription.

Preferential Expression of Antisense Constructs in Embryos

In order to reduce the transcription of a desaturase gene in embryos of *B. napus* or *B. campestris*, constructs may be prepared which allow for production of antisense copies of the desaturase cDNA preferentially in the embryos. Promoter sequences which are desirable to obtain this pattern of expression include, but are not limited to, the ACP, Bce4, and napin 1-2 expression cassettes described in Examples 7, 8, and 9, respectively. It also may be desirable to control the expression of reverse copies of the desaturase cDNA under two different promoters in the same transformed plant to provide for a broader timing of expression of the antisense desaturase DNA. For example, expression from the ACP promoter may begin and end earlier than expression from the napin promoter. Thus, expressing the reverse desaturase from both promoters may result in

the production of the antisense strand of DNA over a longer period of embryo development.

An example of expression of an antisense desaturase gene preferentially in the embryos is provided below. Similar constructs containing the same or a different fragment of the desaturase gene and any of the promoters described above, as well as other promoter regions which may be useful, may also be prepared using gene cloning, insertion, mutation and repair techniques well known to those of ordinary skill in the art.

A. Antisense Desaturase Expression from the ACP Promoter

Construction of pCGN3239 is as follows:

pCGN3235 (Example 12) is digested with *Pvu*II and *Hin*dIII and the *Hin*dIII sticky ends are filled in with Klenow in the presence of 200 µM dNTPs. The 1.2 kb *Pvu*II/*Hin*dIII fragment containing the desaturase coding sequence is gel purified and ligated in the antisense orientation into *Eco*RV-digested pCGN1977 (ACP expression cassette; described in Example 7) to create pCGN3238. The 4.2 kb *Xba*I/*Asp*718 fragment of pCGN3238 containing the antisense desaturase in the ACP cassette is transferred into *Xba*I/*Asp*718-digested pCGN1557 (binary transformation vector; described in Example 7) to create pCGN3239.

B. Antisense Desaturase Expression From The Napin Promoter

Construction of pCGN3240 is as follows: pCGN3235 is digested with *Pvu*II and *Hin*dIII, the sticky ends are blunted, and the resulting fragment is inserted in an anti-sense orientation into pCGN3223 which has been digested with *Sal*I and blunted with Klenow enzyme. The resulting plasmid, pCGN3240 will express an anti-sense desaturase RNA from the napin promoter cassette.

C. Antisense Desaturase Expression From a Dual Promoter Cassette

Construction of pCGN3242 is as follows: An *Asp*718 fragment of pCGN3240 containing the napin 5' and 3' regions surrounding the desaturase sequences is inserted into the

*Asp*718 site of pCGN3239 (a binary vector containing an ACP promoter, antisense desaturase construct) to create pCGN3242.

Constitutive Transcription

A. Binary Vector Construction

1. Construction of pCGP291.

The *Kpn*I, *Bam*HI, and *Xba*I sites of binary vector pCGN1559 (McBride and Summerfelt, *Pl.Mol.Biol.* (1990) *14*: 269-276) are removed by *Asp*718/*Xba*I digestion followed by blunting the ends and recircularization to produce pCGP67. The 1.84 kb *Pst*I/*Hin*dIII fragment of pCGN986 containing the 35S promoter-tml3' cassette is inserted into *Pst*I/*Hin*dIII digested pCGP67 to produce pCGP291.

2. Construction of pCGN986.

The 35S promoter-tml3' expression cassette, pCGN986, contains a cauliflower mosaic virus 35S (CaMV35) promoter and a T-DNA tml 3'-region with multiple restriction sites between them. pCGN986 is derived from another cassette, pCGN206, containing a CaMV35S promoter and a different 3' region, the CaMV region VI 3'-end. The CaMV 35S promoter is cloned as an *Alu*I fragment (bp 7144-7734) (Gardner *et. al., Nucl.Acids Res.* (1981) *9*:2871-2888) into the *Hin*cII site of M13mp7 (Messing, *et. al., Nucl.Acids Res.* (1981) *9*:309-321) to create C614. An *Eco*RI digest of C614 produced the *Eco*RI fragment from C614 containing the 35S promoter which is cloned into the *Eco*RI site of pUC8 (Vieira and Messing, *Gene* (1982) *19*:259) to produce pCGN147.

pCGN148a containing a promoter region, selectable marker (KAN with 2 ATG's) and 3' region, is prepared by digesting pCGN528 with *Bgl*II and inserting the *Bam*HI-*Bgl*II promoter fragment from pCGN147. This fragment is cloned into the *Bgl*II site of pCGN528 so that the *Bgl*II site is proximal to the kanamycin gene of pCGN528.

The shuttle vector used for this construct, pCGN528, is made as follows: pCGN525 is made by digesting a plasmid containing Tn5 which harbors a kanamycin gene (Jorgenson

et. al., Mol. Gen. Genet. (1979) *177*:65) with *Hin*dIII-*Bam*HI and inserting the *Hin*dIII-*Bam*HI fragment containing the kanamycin gene into the *Hin*dIII-*Bam*HI sites in the tetracycline gene of pACYC184 (Chang and Cohen, *J. Bacteriol.* (1978) *134*:1141-1156). pCGN526 was made by inserting the *Bam*HI fragment 19 of pTiA6 (Thomashow *et. al., Cell* (1980) *19*:729-739), modified with *Xho*I linkers inserted into the *Sma*I site, into the *Bam*HI site of pCGN525. pCGN528 is obtained by deleting the small *Xho*I fragment from pCGN526 by digesting with *Xho*I and religating.

pCGN149a is made by cloning the *Bam*HI-kanamycin gene fragment from pMB9KanXXI into the *Bam*HI site of pCGN148a. pMB9KanXXI is a pUC4K variant (Vieira and Messing, *Gene* (1982) *19*:259-268) which has the *Xho*I site missing, but contains a functional kanamycin gene from Tn903 to allow for efficient selection in *Agrobacterium*.

pCGN149a is digested with *Hin*dIII and *Bam*HI and ligated to pUC8 digested with *Hin*dIII and *Bam*HI to produce pCGN169. This removes the Tn903 kanamycin marker. pCGN565 and pCGN169 are both digested with *Hin*dIII and *Pst*I and ligated to form pCGN203, a plasmid containing the CaMV 35S promoter and part of the 5'-end of the Tn5 kanamycin gene (up to the *Pst*I site, Jorgenson *et. al.*, (1979), *supra*). A 3'-regulatory region is added to pCGN203 from pCGN204, an *Eco*RI fragment of CaMV (bp 408-6105) containing the region VI 3' cloned into pUC18 (Yanisch-Perron, *et al., Gene* (1985) *33*:103-119) by digestion with *Hin*dIII and *Pst*I and ligation. The resulting cassette, pCGN206, is the basis for the construction of pCGN986.

The pTiA6 T-DNA tml 3'-sequences are subcloned from the *Bam*19 T-DNA fragment (Thomashow *et al.*, (1980) *supra*) as a *Bam*HI-*Eco*RI fragment (nucleotides 9062 to 12,823, numbering as in Barker *et al., Plant Mol. Biol.* (1982) *2*:335-350) and combined with the pACYC184 (Chang and Cohen (1978), *supra*) origin of replication as an *Eco*RI-*Hin*dIII fragment and a gentamycin resistance marker (from plasmid

pLB41), obtained from D. Figurski) as a *Bam*HI-*Hin*dIII fragment to produce pCGN417.

The unique *Sma*I site of pCGN417 (nucleotide 11,207 of the *Bam*19 fragment) is changed to a *Sac*I site using linkers and the *Bam*HI-*Sac*I fragment is subcloned into pCGN565 to give pCGN971. The *Bam*HI site of pCGN971 is changed to an *Eco*RI site using linkers. The resulting *Eco*RI-*Sac*I fragment containing the tml 3' regulatory sequences is joined to pCGN206 by digestion with *Eco*RI and *Sac*I to give pCGN975. The small part of the Tn5 kanamycin resistance gene is deleted from the 3'-end of the CaMV 35S promoter by digestion with *Sal*I and *Bgl*II, blunting the ends and ligation with *Sal*I linkers. The final expression cassette pCGN986 contains the CaMV 35S promoter followed by two *Sal*I sites, an *Xba*I site, *Bam*HI, *Sma*I, *Kpn*I and the tml 3' region (nucleotides 11207-9023 of the T-DNA).

B. <u>Insertion of Desaturase Sequence</u>

The 1.6 kb *XbaI* fragment from pCGN3235 containing the desaturase cDNA is inserted in the antisense orientation into the *Xba*I site of pCGP291 to produce pCGN3234.

Plant Transformation

The binary vectors containing the expression cassette and the desaturase gene are transformed into *Agrobacterium tumefaciens* strain EHA101 (Hood, *et al.*, *J. Bacteriol.* (1986) *168*:1291-1301) as per the method of Holsters, *et al.*, *Mol. Gen. Genet.* (1978) *163*:181-187. Transformed *B. napus* and/or *Brassica campestris* plants are obtained as described in Example 10.

Analysis of Transgenic Plants

A. <u>Analysis of pCGN3242 Transformed *Brassica campestris* cv. Tobin Plants</u>

Due to the self-incompatibility of *Brassica campestris* cv. Tobin, individual transgenic plants are pollinated using non-transformed Tobin pollen. Because of this, the T2 seeds of a transgenic plant containing the antisense desaturase at one locus would be expected to segregate in a

1:1 ratio of transformed to non-transformed seed. The oil composition of ten individual seeds collected at 26 days post-anthesis from several pCGN3242-transformed plants and one non-transformed control was analyzed by gas chromatography according to the method of Browse, *et al.*, *Anal. Biochem.* (1986) *152*:141-145. One transformant, 3242-T-1, exhibits an oil composition that differed distinctly from controls on preliminary analysis. The control Tobin seeds contained an average of 1.8% 18:0 (range 1.5% - 2.0%) and 52.9% 18:1 (range 48.2% - 57.1%). T2 seeds of 3242-T-1 segregated into two distinct classes. Five seeds contained levels of 18:0 ranging from 1.3% to 1.9% and levels of 18:1 ranging from 42.2% to 58.3%. The other five seeds contained from 22.9% to 26.3% 18:0 and from 19.9% to 26.1% 18:1.

B. Analysis of pCGN3234 Transformed Plants

Some abnormalities have been observed in some transgenic *Brassica napus* cv. Delta and Bingo and *Brassica campestris* cv. Tobin plants containing pCGN3234. These effects could be due to the constitutive expression of antisense desaturase RNA from the 35S promoter or could be due to the transformation/tissue culture regime the plants have been subjected to.

The above results demonstrate the ability to obtain plant Δ-9 desaturases, isolate DNA sequences which encode desaturase activity and manipulate them. In this way, the production of transcription cassettes, including expression cassettes can be produced which allow for production, including specially differentiated cell production of the desired product. A purified *C. tinctorius* desaturase is provided and used to obtain nucleic acid sequences of *C. tinctorius* desaturase. Other plant desaturase sequences are provided such as *R. cummunis*, *B. campestris*, and *S. chinensis*. These sequences as well as desaturase sequences obtained from them may be used to obtain additional desaturease, and so on. And, as described in the application modification of oil composition may b achieved.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claim.

What is claimed is:

1. A recombinant DNA construct comprising a sequence encoding at least a portion of a plant desaturase, said desaturase when mature having activity toward an unsaturated fatty acid substrate.

2. The construct of Claim 1 encoding a biologically active plant desaturase.

3. The construct of Claim 1 wherein said sequence encodes a precursor desaturase.

4. The construct of Claim 1 wherein said sequence encodes a mature desaturase.

5. The construct of Claim 1 wherein said sequence encodes a transit peptide.

6. The construct of Claim 1 comprising a cDNA sequence.

7. The construct of Claim 1 wherein said sequence is joined to a second nucleic acid sequence which is not naturally joined to said first sequence.

8. The construct of Claim 1 comprising, in the 5' to 3' direction of transcription, a transcriptional regulatory region functional in a host cell and said sequence.

9. The construct of Claim 8 further comprising, a translational regulatory region immediately 5' to said sequence and a transcriptional/translational termination regulatory region 3' to said sequence, wherein said regulatory regions are functional in said host cell.

10. The construct of Claim 8 wherein said sequence is a sense sequence.

11. The construct of Claim 8 wherein said sequence is an anti-sense sequence.

12. The construct of Claim 8 wherein said host cell is a plant cell.

13. The construct of Claim 12 wherein said transcriptional initiation region is obtained from a gene preferentially expressed in plant seed tissue during lipid accumulation.

14. The construct of Claim 13 wherein said transcriptional initiation region is selected from the

regulatory region 5' upstream to a structural gene of the group consisting of any one of Bce4, seed ACP Bcg 4-4 and napin 1-2.

15. The construct of Claim 9 wherein said transcriptional termination region is a plant desaturase termination region.

16. The construct of Claim 1 wherein said plant desaturase is a Δ-9 desaturase.

17. The construct of Claim 1 wherein said sequence is obtainable from any one of *C. tinctorius, R. communinis* and *B. campestris*.

18. A method of modifying fatty acid composition in a plant host cell from a given percentage of fatty acid saturation to a different percentage of fatty acid saturation comprising

growing a host plant cell having integrated into its genome a recombinant DNA sequence encoding a plant desaturase under the control of regulatory elements functional in said plant cell during lipid accumulation under conditions which will promote the activity of said regulatory elements.

19. The method of Claim 18 wherein the overexpression of plant desaturase is obtained.

20. The method of Claim 18 wherein the decrease of endogenous plant desaturase is obtained.

21. The method of Claim 18 wherein said regulatory elements function preferentially in plant seed.

22. The method of Claim 20 wherein the percentage of long chain unsaturated fatty acids is increased.

23. A plant cell having a modified level of saturated fatty acids produced according to the method of any one of Claims 18-22.

24. The plant cell of Claim 23 wherein said cell is a *Brassica* plant cell.

25. The plant cell of Claim 23 wherein said cell is *in vivo*.

26. The plant cell of Claim 23 wherein said cell is an oilseed embryo plant cell.

27. A plant seed having a modified level of saturated fatty acids as compared to a seed of said plant having a native level of saturated fatty acids produced according to a method comprising

growing a plant, having integrated into the genome of embryo cells a recombinant DNA sequence encoding a plant desaturase under the control of regulatory elements functional in seed during lipid accumulation, to produce seed under conditions which will promote the activity of said regulatory elements, and

harvesting said seed.

28. The seed of Claim 27 wherein said plant is *Brassica napus*.

29. The seed of claim 27 wherein said seed is an oilseed.

30. The seed of Claim 27 wherein said plant desaturase is a Δ-9 desaturase.

31. A plant seed oil of a plant having an endogenous level of saturated fatty acids comprising a plant seed oil having a modified level of saturated fatty acids.

32. The oil of Claim 31 comprising a *Brassica napus* oil.

33. A plant seed oil separated from an seed produced according to any one of Claims 27-30.

34. A host cell comprising a plant desaturase encoding sequence of any one of Claims 1-17.

35. The cell of Claim 34 wherein said cell is a plant cell.

36. The cell of Claim 35 wherein said plant cell is *in vivo*.

37. The cell of Claim 35 wherein said plant cell is a *Brassica* plant cell.

38. A transgenic host cell comprising an expressed plant desaturase.

39. The cell of Claim 38 wherein said host cell is a plant cell.

40. The cell of Claim 38 wherein said plant desaturase is a Δ-9 desaturase.

41. A method of producing a plant desaturase in a host cell or progeny thereof comprising

growing a host cell or progeny thereof comprising a construct of any one of Claims 1-10 and 12-17 under conditions which will permit the production of said plant desaturase.

42. The method of Claim 41 wherein said host cell is a plant cell and said construct is integrated into the genome of said plant cell.

43. The method of Claim 42 wherein said plant cell is *in vivo*.

44. A host cell comprising a plant desaturase produced according to Claim 41.

45. The cell of Claim 45 wherein said host cell is a plant host cell and said construct is integrated into the genome of said plant cell.

F1: ASTLGSSTPKVDNAKKPFQPPREVHVQVTH$^{S}_{X}$MPPQKIEIFKSIEG$^{W}_{R}$AEQNILV$^{H}_{F}$LKPVEKCWQ

F2: DFLPDPA$^{S}_{T}$EGFDEQVKELRARAKEIPDDYFVVLVGDMITEEALPTYQTMLNTLDGV

F3: DETGASLTPWAVWT

F4: DLLHTYLYLSGRV

F5: DMRQIQKTIQYLI

F6: TENSPYLGFIYTSFQER

F7: DV$^{K}_{F}$LAQI$^{C}_{Q}$GTIASDEKRHETAYTKIVEKLFEIDPDGTVLAFADMMRKKI$^{S}_{T}$MPAHLMY

F8: DNLF

F9: dvFlAV$^{A}_{I}$QRL$^{G}_{I}$VYTAK

F10: DYADILEFLVGRWK

F11: VADLTGLSGEGRKA$^{Q}_{G}$DYVCGLPPRIRRLEERAQGRAKEGPVVPFSWIFDRQVKL

FIGURE 1

```
                                                         HindIII
                                                         |
  1 GCTCACTTGTGTGGTGGAGGAGAAAAACAGAACTCACAAAAAGCTTTGCGACTGCCAAGAACAACAACA   69
                                                         42

 70 ACAACAAGATCAAGAAGAAGAAGAAGAAGATCAAAAATGGCTCTTCGAATCACTCCAGTGACCTTGCAA  138
                                         METAlaLeuArgIleThrProValThrLeuGln

          EcoRV                                          BglII        NcoI
          |                                              |            |
139 TCGGAGAGATATCGTTCGTTTTCGTTTCCTAAGAAGGCTAATCTCAGATCTCCCAAATTCGCCATGGCC  207
    SerGluArgTyrArgSerPheSerPheProLysLysAlaAsnLeuArgSerProLysPheAlaMETAla
          149                                            185          201

                                 HindII
                                 |
208 TCCACCCTCGGATCATCCACACCGAAGGTTGACAATGCCAAGAAGCCTTTTCAACCTCCACGAGAGGTT  276
    SerThrLeuGlySerSerThrProLysValAspAsnAlaLysLysProPheGlnProProArgGluVal
                                 238

277 CATGTTCAGGTGACGCACTCCATGCCACCACAGAAGATAGAGATTTTCAAATCCATCGAGGGTTGGGCT  345
    HisValGlnValThrHisSerMETProProGlnLysIleGluIlePheLysSerIleGluGlyTrpAla

346 GAGCAGAACATATTGGTTCACCTAAAGCCAGTGGAGAAATGTTGGCAAGCACAGGATTTCTTGCCGGAC  414
    GluGlnAsnIleLeuValHisLeuLysProValGluLysCysTrpGlnAlaGlnAspPheLeuProAsp
```

FIGURE 2

Page 1 of 4

```
415 CCTGCATCTGAAGGATTTGATGAACAAGTCAAGGAACTAAGGGCAAGAGCAAAGGAGATTCCTGATGAT  483
    ProAlaSerGluGlyPheAspGluGlnValLysGluLeuArgAlaArgAlaLysGluIleProAspAsp

484 TACTTTGTTGTTTTGGTTGGAGATATGATTACAGAGGAAGCCCTACCTACTTACCAAACAATGCTTAAT  552
    TyrPheValValLeuValGlyAspMETIleThrGluGluAlaLeuProThrTyrGlnThrMETLeuAsn

553 ACCCTAGATGGTGTACGTGATGAGACTGGGGCTAGCCTTACGCCTTGGGCTGTCTGGACTAGGGCTTGG  621
    ThrLeuAspGlyValArgAspGluThrGlyAlaSerLeuThrProTrpAlaValTrpThrArgAlaTrp

        PvuII                                                           AccI
        |                                                               |
622 ACAGCTGAAGAGAACAGGCATGGCGATCTTCTCCACACCTATCTCTACCTTTCTGGGCGGTAGACATG  690
    ThrAlaGluGluAsnArgHisGlyAspLeuLeuHisThrTyrLeuTyrLeuSerGlyArgValAspMET
        626                                                             684

                                                    BamHI
                                                    |
691 AGGCAGATACAGAAGACAATTCAGTATCTCATTGGGTCAGGAATGGATCCTCGTACCGAAAACAGCCCC  759
    ArgGlnIleGlnLysThrIleGlnTyrLeuIleGlySerGlyMETAspProArgThrGluAsnSerPro
                                                    736

760 TACCTTGGGTTCATCTACACATCGTTTCAAGAGCGTGCCACATTTGTTTCTCACGGAAACACCGCCAGG  828
    TyrLeuGlyPheIleTyrThrSerPheGlnGluArgAlaThrPheValSerHisGlyAsnThrAlaArg
```

FIGURE 2
Page 2 of 4

```
     SphI
     |
829  CATGCAAAGGATCATGGGGACGTGAAACTGGCGCAAATTTGTGGTACAATCGCGTCTGACGAAAAGCGT  897
     HisAlaLysAspHisGlyAspValLysLeuAlaGlnIleCysGlyThrIleAlaSerAspGluLysArg
     833

                                                        ClaI
                                                        |
898  CACGAGACCGCTTATACAAAGATAGTCGAAAAGCTATTCGAGATCGATCCTGATGGCACCGTTCTTGCT  966
     HisGluThrAlaTyrThrLysIleValGluLysLeuPheGluIleAspProAspGlyThrValLeuAla
                                                        942

                          BglII
                          |
967  TTTGCCGACATGATGAGGAAAAAGATCTCGATGCCCGCACACTTGATGTACGATGGGCGTGATGACAAC 1035
     PheAlaAspMETMETArgLysLysIleSerMETProAlaHisLeuMETTyrAspGlyArgAspAspAsn
                          990

                                                     AccI
                                                     |
1036 CTCTTCGAACATTTCTCGGCGGTTGCCCAAAGACTCGGCGTCTACACCGCCAAAGACTACGCCGACATA 1104
     LeuPheGluHisPheSerAlaValAlaGlnArgLeuGlyValTyrThrAlaLysAspTyrAlaAspIle
                                                     1077

1105 CTGGAATTTCTGGTCGGGCGGTGGAAAGTGGCGGATTTGACCGGCCTATCTGGTGAAGGGCGTAAAGCG 1173
     LeuGluPheLeuValGlyArgTrpLysValAlaAspLeuThrGlyLeuSerGlyGluGlyArgLysAla
```

FIGURE 2
Page 3 of 4

```
                                                                  SacI
                                                                  |
1174 CAAGATTATGTTTGCGGGTTGCCACCAAGAATCAGAAGGCTGGAGGAGAGAGCTCAAGGGCGAGCAAAG 1242
     GlnAspTyrValCysGlyLeuProProArgIleArgArgLeuGluGluArgAlaGlnGlyArgAlaLys
                                                                  1228

                           PvuII
                           |
1243 GAAGGACCTGTTGTTCCATTCAGCTGGATTTTCGATAGACAGGTGAAGCTGTGAAGAAAAAAAAAACGA 1311
     GluGlyProValValProPheSerTrpIlePheAspArgGlnValLysLeu
                           1266

1312 GCAGTGAGTTCGGTTTCTGTTGGCTTATTGGGTAGAGGTTAAAACCTATTTAGATGTCTGTTTCGTGT 1380

1381 AATGTGGTTTTTTTCTTCTAATCTTGAATCTGGTATTGTGTCGTTGAGTTCGCGTGTGTGTAAACTTG 1449

1450 TGTGGCTGTGGACATATTATAGAACTCGTTATGCCAATTTTGATGACGGTGGTTATCGTCTCCCCTGGT 1518

1519 GTTTTTTTATTGTTT 1533
```

FIGURE 2
Page 4 of 4

```
  1 AAAAGAAAAAGGTAAGAAAAAAAACAATGGCTCTCAAGCTCAATCCTTTCCTTTCTCAAACCCAAAAGT  69
                                METAlaLeuLysLeuAsnProPheLeuSerGlnThrGlnLysL

                                           BglII
                                           |
 70 TACCTTCTTTCGCTCTTCCACCAATGGCCAGTACCAGATCTCCTAAGTTCTACATGGCCTCTACCCTCA 138
    euProSerPheAlaLeuProProMETAlaSerThrArgSerProLysPheTyrMETAlaSerThrLeuL

139 AGTCTGGTTCTAAGGAAGTTGAGAATCTCAAGAAGCCTTTCATGCCTCCTCGGGAGGTACATGTTCAGG 207
    ysSerGlySerLysGluValGluAsnLeuLysLysProPheMETProProArgGluValHisValGlnV

208 TTACCCATTCTATTGCCA 225
    alThrHisSerIleAla
```

FIGURE 3A

```
AAAAGAAAAA GGTAAGAAAA AAAACA ATG GCT CTC AAG CTC AAT CCT TTC CTT TCT       56
                             MET Ala Leu Lys Leu Asn Pro Phe Leu Ser

CAA ACC CAA AAG TTA CCT TCT TTC GCT CTT CCA CCA ATG GCC AGT ACC AGA TCT   110
Gln Thr Gln Lys Leu Pro Ser Phe Ala Leu Pro Pro MET Ala Ser Thr Arg Ser

CCT AAG TTC TAC ATG GCC TCT ACC CTC AAG TCT GGT TCT AAG GAA GTT GAG AAT   164
Pro Lys Phe Tyr MET Ala Ser Thr Leu Lys Ser Gly Ser Lys Glu Val Glu Asn

CTC AAG AAG CCT TTC ATG CCT CCT CGG GAG GTA CAT GTT CAG GTT ACC CAT TCT   218
Leu Lys Lys Pro Phe MET Pro Pro Arg Glu Val His Val Gln Val Thr His Ser

ATG CCA CCC CAA AAG ATT GAG ATC TTT AAA TCC CTA GAC AAT TGG GCT GAG GAG   272
MET Pro Pro Gln Lys Ile Glu Ile Phe Lys Ser Leu Asp Asn Trp Ala Glu Glu

AAC ATT CTG GTT CAT CTG AAG CCA GTT GAG AAA TGT TGG CAA CCG CAG GAT TTT   326
Asn Ile Leu Val His Leu Lys Pro Val Glu Lys Cys Trp Gln Pro Gln Asp Phe

TTG CCA GAT CCC GCC TCT GAT GGA TTT GAT GAG CAA GTC AGG GAA CTC AGG GAG   380
Leu Pro Asp Pro Ala Ser Asp Gly Phe Asp Glu Gln Val Arg Glu Leu Arg Glu

AGA GCA AAG GAG ATT CCT GAT GAT TAT TTT GTT GTT TTG GTT GGA GAC ATG ATA   434
Arg Ala Lys Glu Ile Pro Asp Asp Tyr Phe Val Val Leu Val Gly Asp MET Ile

ACG GAA GAA GCC CTT CCC ACT TAT CAA ACA ATG CTG AAT ACC TTG GAT GGA GTT   488
Thr Glu Glu Ala Leu Pro Thr Tyr Gln Thr MET Leu Asn Thr Leu Asp Gly Val
```

FIGURE 3B
Page 1 of 3

```
CGG GAT GAA ACA GGT GCA AGT CCT ACT TCT TGG GCA ATT TGG ACA AGG GCA TGG     542
Arg Asp Glu Thr Gly Ala Ser Pro Thr Ser Trp Ala Ile Trp Thr Arg Ala Trp

ACT GCG GAA GAG AAT AGA CAT GGT GAC CTC CTC AAT AAG TAT CTC TAC CTA TCT     596
Thr Ala Glu Glu Asn Arg His Gly Asp Leu Leu Asn Lys Tyr Leu Tyr Leu Ser

GGA CGA GTG GAC ATG AGG CAA ATT GAG AAG ACA ATT CAA TAT TTG ATT GGT TCA     650
Gly Arg Val Asp MET Arg Gln Ile Glu Lys Thr Ile Gln Tyr Leu Ile Gly Ser

GGA ATG GAT CCA CGG ACA GAA AAC AGT CCA TAC CTT GGG TTC ATC TAT ACA TCA     704
Gly MET Asp Pro Arg Thr Glu Asn Ser Pro Tyr Leu Gly Phe Ile Tyr Thr Ser

TTC CAG GAA AGG GCA ACC TTC ATT TCT CAT GGG AAC ACT GCC CGA CAA GCC AAA     758
Phe Gln Glu Arg Ala Thr Phe Ile Ser His Gly Asn Thr Ala Arg Gln Ala Lys

GAG CAT GGA GAC ATA AAG TTG GCT CAA ATA TGT GGT ACA ATT GCT GCA GAT GAG     812
Glu His Gly Asp Ile Lys Leu Ala Gln Ile Cys Gly Thr Ile Ala Ala Asp Glu

AAG CGC CAT GAG ACA GCC TAC ACA AAG ATA GTG GAA AAA CTC TTT GAG ATT GAT     866
Lys Arg His Glu Thr Ala Tyr Thr Lys Ile Val Glu Lys Leu Phe Glu Ile Asp

CCT GAT GGA ACT GTT TTG GCT TTT GCT GAT ATG ATG AGA AAG AAA ATT TCT ATG     920
Pro Asp Gly Thr Val Leu Ala Phe Ala Asp MET MET Arg Lys Lys Ile Ser MET

CCT GCA CAC TTG ATG TAT GAT GGC CGA GAT GAT AAT CTT TTT GAC CAC TTT TCA     974
Pro Ala His Leu MET Tyr Asp Gly Arg Asp Asp Asn Leu Phe Asp His Phe Ser
```

FIGURE 3B
Page 2 of 3

```
GCT GTT GCG CAG CGT CTT GGA GTC TAC ACA GCA AAG GAT TAT GCA GAT ATA TTG            1028
Ala Val Ala Gln Arg Leu Gly Val Tyr Thr Ala Lys Asp Tyr Ala Asp Ile Leu

GAG TTC TTG GTG GGC AGA TGG AAG GTG GAT AAA CTA ACG GGC CTT TCA GCT GAG            1082
Glu Phe Leu Val Gly Arg Trp Lys Val Asp Lys Leu Thr Gly Leu Ser Ala Glu

GGA CAA AAG GCT CAG GAC TAT GTT TGT CGG TTA CCT CCA AGA ATT AGA AGG CTG            1136
Gly Gln Lys Ala Gln Asp Tyr Val Cys Arg Leu Pro Pro Arg Ile Arg Arg Leu

GAA GAG AGA GCT CAA GGA AGG GCA AAG GAA GCA CCC ACC ATG CCT TTC AGC TGG            1190
Glu Glu Arg Ala Gln Gly Arg Ala Lys Glu Ala Pro Thr MET Pro Phe Ser Trp

ATT TTC GAT AGG CAA GTG AAG CTG TAGGTGGCTA AAGTGCAGGA CGAAACCGAA ATGGTTAGTT    1254
Ile Phe Asp Arg Gln Val Lys Leu

TCACTCTTTT TCATGCCCAT CCCTGCAGAA TCAGAAGTAG AGGTAGAATT TTGTAGTTGC TTTTTTATTA  1324

CAAGTCCAGT TTAGTTTAAG GTCTGTGGAA GGGAGTTAGT TGAGGAGTGA ATTTAGTAAG TTGTAGATAC  1394

AGTTGTTTCT TGTGTTGTCA TGAGTATGCT GATAGAGAGC AGCTGTAGTT TTGTTGTTGT GTTCTTTTAT  1464

ATGGTCTCTT GTATGAGTTT CTTTTCTTTC CTTTTCTTCT TTCCTTTCCT CTCTCTCTCT CTCTCTCTCT  1534

CTCTTTTTCT CTTATCCCAA GTGTCTCAAG TATAATAAGC AAACGATCCA TGTGGCAATT TTGATGATGG  1604

TGATCAGTCT CACAACTTGA TCTTTTGTCT TCTATTGGAA ACACAGCCTG CTTGTTTGAA AAAA        1668
```

FIGURE 3B
Page 3 of 3

pCGN3235

```
 1 TGAGAGATAGTGTGAGAGCATTAGCCTTAGAGAGAGAGAGAGAGAGCTTGTGTCTGAAAGAATCCACAA  69

              HindIII
              |
70 ATGGCATTGAAGCTTAACCCTTTGGCATCTCAGCCTTACAACTTCCCT 117
   METAlaLeuLysLeuAsnProLeuAlaSerGlnProTyrAsnPhePro
```

FIGURE 4A

pCGN3236

```
                                                PstI
                                                 |
  1 ACTTCATGGGCTATTTGGACAAGAGCTTGGACTGCAGAAGAGAACCGACACGGTGATCTTCTCAATAAG   69
    ThrSerTrpAlaIleTrpThrArgAlaTrpThrAlaGluGluAsnArgHisGlyAspLeuLeuAsnLys

 70 TATCTTTACTTGTCTGGACGTGTTGACATGAGGCAGATTGAAAAGACCATTCAGTACTTGATTGGTTCT 138
    TyrLeuTyrLeuSerGlyArgValAspMETArgGlnIleGluLysThrIleGlnTyrLeuIleGlySer

          BamHI
           |
139 GGAATGGATCCTAGAACAGAGAACAATCCTTACCTCGG 176
    GlyMETAspProArgThrGluAsnAsnProTyrLeuAla
```

FIGURE 4B

12/42

pCGN3235

```
TGAGAGATAG TGTGAGAGCA TTAGCCTTAG AGAGAGAGAG AGAGAGCTTG TGTCTGAAAG AATCCACAA

ATG GCA TTG AAG CTT AAC CCT TTG GCA TCT CAG CCT TAC AAC TTC CCT TCC TCG
MET Ala Leu Lys Leu Asn Pro Leu Ala Ser Gln Pro Tyr Asn Phe Pro Ser Ser

GCT CGT CCG CCA ATC TCT ACT TTC AGA TCT CCC AAG TTC CTC TGC CTC GCT TCT
Ala Arg Pro Pro Ile Ser Thr Phe Arg Ser Pro Lys Phe Leu Cys Leu Ala Ser

TCT TCT CCC GCT CTC AGC TCC AAG GAG GTT GAG AGT TTG AAG AAG CCA TTC ACA
Ser Ser Pro Ala Leu Ser Ser Lys Glu Val Glu Ser Leu Lys Lys Pro Phe Thr

CCA CCT AAG GAA GTG CAC GTT CAA GTC CTG CAT TCC ATG CCA CCC CAG AAG ATC
Pro Pro Lys Glu Val His Val Gln Val Leu His Ser MET Pro Pro Gln Lys Ile

GAG ATC TTC AAA TCC ATG GAA GAC TGG GCC GAG CAG AAC CTT CTA ACT CAG CTC
Glu Ile Phe Lys Ser MET Glu Asp Trp Ala Glu Gln Asn Leu Leu Thr Gln Leu

AAA GAC GTG GAG AAG TCG TGG CAG CCC CAG GAC TTC TTA CCC GAC CCT GCA TCC
Lys Asp Val Glu Lys Ser Trp Gln Pro Gln Asp Phe Leu Pro Asp Pro Ala Ser

GAT GGG TTC GAA GAT CAG GTT AGA GAG CTA AGA GAG AGG GCA AGA GAG CTC CCT
Asp Gly Phe Glu Asp Gln Val Arg Glu Leu Arg Glu Arg Ala Arg Glu Leu Pro

GAT GAT TAC TTC GTT GTT CTG GTG GGA GAC ATG ATC ACG GAA GAG GCG CTT CCG
Asp Asp Tyr Phe Val Val Leu Val Gly Asp MET Ile Thr Glu Glu Ala Leu Pro
```

FIGURE 4C
Page 1 of 3

ACC TAT CAA ACC ATG TTG AAC ACT TTG GAT GGA GTG AGG GAT GAA ACT GGC GCT
Thr Tyr Gln Thr MET Leu Asn Thr Leu Asp Gly Val Arg Asp Glu Thr Gly Ala

AGC CCC ACT TCA TGG GCT ATT TGG ACA AGA GCT TGG ACT GCA GAA GAG AAC CGA
Ser Pro Thr Ser Trp Ala Ile Trp Thr Arg Ala Trp Thr Ala Glu Glu Asn Arg

CAC GGT GAT CTT CTC AAT AAG TAT CTT TAC TTG TCT GGA CGT GTT GAC ATG AGG
His Gly Asp Leu Leu Asn Lys Tyr Leu Tyr Leu Ser Gly Arg Val Asp MET Arg

CAG ATT GAA AAG ACC ATT CAG TAC TTG ATT GGT TCT GGA ATG GAT CCT AGA ACA
Gln Ile Glu Lys Thr Ile Gln Tyr Leu Ile Gly Ser Gly MET Asp Pro Arg Thr

GAG AAC AAT CCT TAC CTC GGC TTC ATC TAC ACT TCA TTC CAA GAA AGA GCC ACC
Glu Asn Asn Pro Tyr Leu Gly Phe Ile Tyr Thr Ser Phe Gln Glu Arg Ala Thr

TTC ATC TCT CAC GGA AAC ACA GCT CGC CAA GCC AAA GAG CAC GGA GAC CTC AAG
Phe Ile Ser His Gly Asn Thr Ala Arg Gln Ala Lys Glu His Gly Asp Leu Lys

CTA GCC CAA ATC TGC GGC ACA ATA GCT GCA GAC GAG AAG CGT CAT GAG ACA GCT
Leu Ala Gln Ile Cys Gly Thr Ile Ala Ala Asp Glu Lys Arg His Glu Thr Ala

TAC ACC AAG ATA GTT GAG AAG CTC TTT GAG ATT GAT CCT GAT GGT ACT GTG ATG
Tyr Thr Lys Ile Val Glu Lys Leu Phe Glu Ile Asp Pro Asp Gly Thr Val MET

GCG TTT GCA GAC ATG ATG AGG AAG AAA ATC TCG ATG CCT GCT CAC TTG ATG TAC
Ala Phe Ala Asp MET MET Arg Lys Lys Ile Ser MET Pro Ala His Leu MET Tyr

FIGURE 4C
Page 2 of 3

```
GAT GGG CGG GAT GAA AGC CTC TTT GAC AAC TTC TCT TCT GTT GCT CAG AGG CTC
Asp Gly Arg Asp Glu Ser Leu Phe Asp Asn Phe Ser Ser Val Ala Gln Arg Leu

GGT GTT TAC ACT GCC AAA GAC TAT GCG GAC ATT CTT GAG TTT TTG GTT GGG AGG
Gly Val Tyr Thr Ala Lys Asp Tyr Ala Asp Ile Leu Glu Phe Leu Val Gly Arg

TGG AAG ATT GAG AGC TTG ACC GGG CTT TCA GGT GAA GGA AAC AAA GCG CAA GAG
Trp Lys Ile Glu Ser Leu Thr Gly Leu Ser Gly Glu Gly Asn Lys Ala Gln Glu

TAC TTG TGT GGG TTG ACT CCA AGA ATC AGG AGG TTG GAT GAG AGA GCT CAA GCA
Tyr Leu Cys Gly Leu Thr Pro Arg Ile Arg Arg Leu Asp Glu Arg Ala Gln Ala

AGA GCC AAG AAA GGA CCC AAG GTT CCT TTC AGC TGG ATA CAT GAC AGA GAA GTG
Arg Ala Lys Lys Gly Pro Lys Val Pro Phe Ser Trp Ile His Asp Arg Glu Val

CAG CTC TAA AAAGGAA CAAAGCTATG AAACCTTTTC ACTCTCCGTC GTCCCTCATT TGATCTATCT
Gln Leu  *

GCTCTTGAAA TTGGTGTAGA TTACTATGGT TTGTGATATT GTTCGTGGGT CTAGTTACAA AGTTGAGAAG

CAGTGATTTA GTAGCTTTGT TGTTTCCAGT CTTTAAATGT TTTTGTGTTT GGTCCTTTTA GTAAACTTGT

TGTAGTTAAA TCAGTTGAAC TGTTTGGTCT GT
```

FIGURE 4C
Page 3 of 3

```
GAT GCC AAA ANG CCT CAC ATG CCT CCT AGA GAA GCT CAT GTG CAA AAG    48
Asp Ala Lys Xaa Pro His MET Pro Pro Arg Glu Ala His Val Gln Lys
 1               5                  10                  15

ACC CAT TCA ATK CCG CCT CAA AAG ATT GAG ATT TTC AAA TCC TTG GAG    96
Thr His Ser Xaa Pro Pro Gln Lys Ile Glu Ile Phe Lys Ser Leu Glu
            20                  25                  30

GGT TGG GCT GAG GAG AAT GTC TTG GTG CAT CTT AAA CCT GTG GAG AA    143
Gly Trp Ala Glu Glu Asn Val Leu Val His Leu Lys Pro Val Glu
        35                  40                  45
```

FIGURE 5

16/42

```
Amino Acid
Sequence From
Fragment F2   K   E   I   P   D   D   Y  FVVLVGDMITEEALPTY  Q   T   M   L   N   T

              AAA GAA AUU CCN GAU GAU UAU                   CAA ACN AUG CUN AAU AC/N
               G   G   C       C   C   C                     G               C

                       A
                                      Forward Primers:
5'GCTAAGCTT AAP GAP ATQ CCA GAQ GAQ TA3'  Desat 13-1
                      A CCG               Desat 13-2
                        CCC               Desat 13-3
                        CCT               Desat 13-4

                                      Reverse Primers:
                                      (complements)

                                       Desat 13-5a   3' GTQ TGN TAC GAN TTP TGCTTAAGCGA 5'
                                       Desat 13-6a                  AAQ
```

Oligonucleotides

P = A or G
Q = T or C
N = A, C, T or C

FIGURE 6

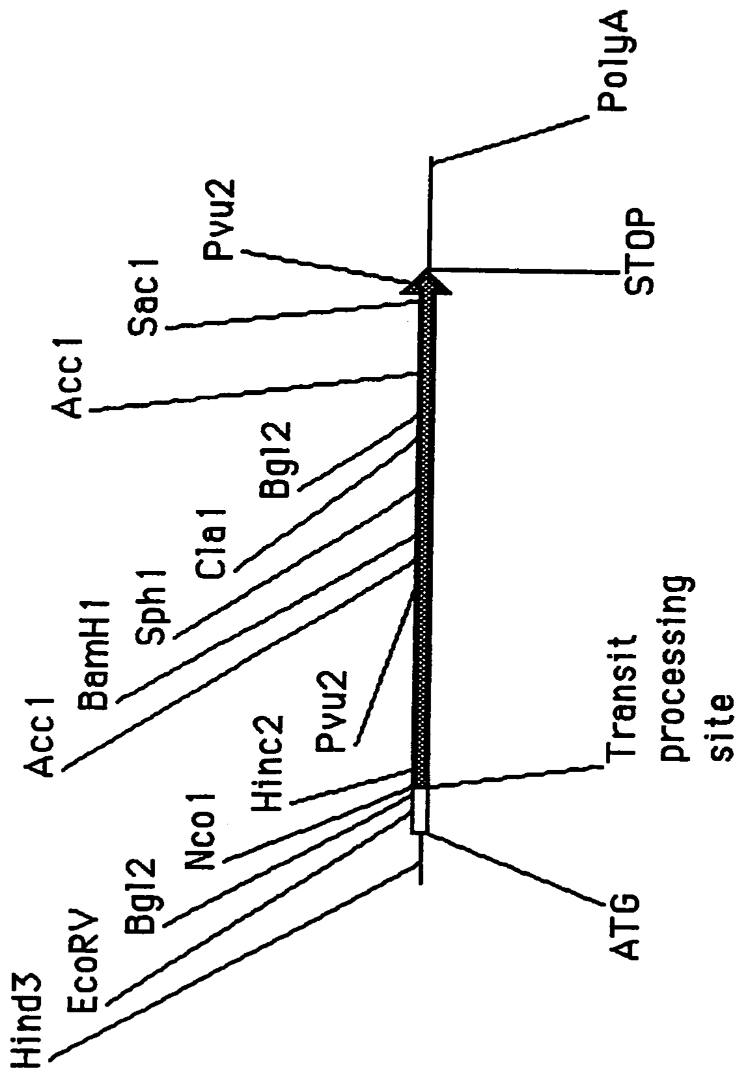

Hind3
EcoRV
Bgl2
Nco1
Hinc2
Pvu2
Acc1
BamH1
Sph1
Cla1
Bgl2
Acc1
Sac1
Pvu2
PolyA
STOP
Transit processing site
ATG

FIGURE 7A

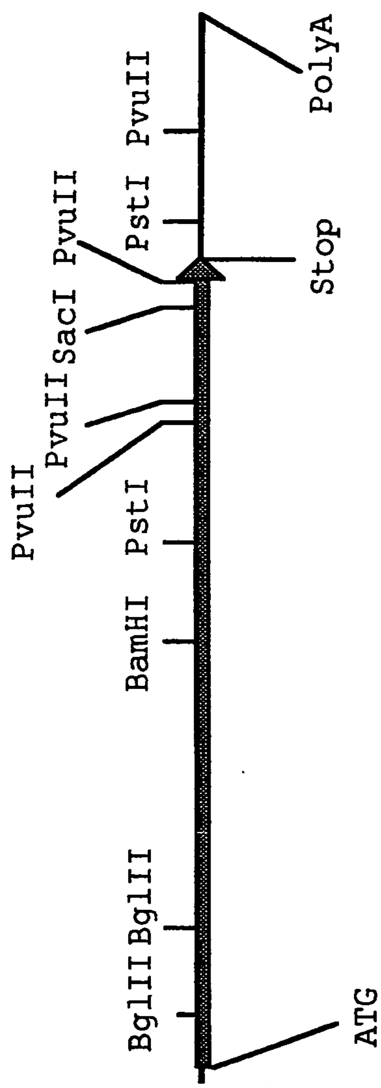

BglIII
BglIII
BamHI
PstI
PvuII
PvuII
SacI
PvuII
PstI
PvuII
PolyA
Stop
ATG

FIGURE 7B

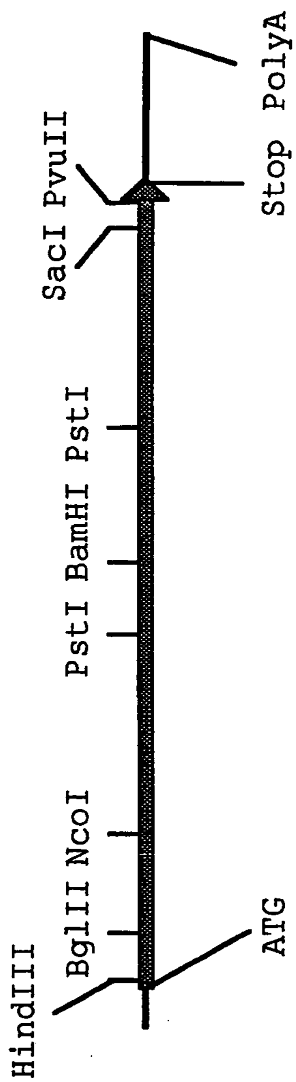

HindIII
BglII
NcoI
PstI
BamHI
PstI
SacI
PvuII
Stop
PolyA
ATG

FIGURE 7C

20/42

```
TCTAGAATTC TCTAATTACG TCTGTTTGTT CTATTTTTTA TATGATATCA AATATTCGTC ATAAATATAT   70
GGTTTAAGAT GCCAAAAAAT TATTTACTTG GTGAATATAA TACGTTAAAT ATTAGAAATA CATCATTTAG  140
TTAAATAAAT AACCAAAAAC CAAAAATTCA TATCCGCGCT GGCGCGCGGT CAGGGTCTCG TTAGTTTTAA  210
AATCAATGCA GTTTACAATT AATTTCCAGC TGAAAATAAG TATAATTTGT ATTGAAATTA TAAAGTGACA  280
TTTTTTGTGT AACAAATATT TTGTGTAACA AGAATTAAAA AAAAAAACAG AAAATACTCA GCTTTTTTAA  350
TAATAAAAAA AATTAATTGA GTTAGAAAAT TGTTGTACCA ATAACAAAAG ATTTATATGG AATTATAAAA  420
TCAACACACC AATAACACAA GACTTTTTAA AAATTTAAGA ATAATATAAG CAATAACAAT AGAATCTTCA  490
AATTCTTCAA ATCCTTAAAA ATCAATCTCC CACTATTAAT CCCCCTTAGT TTTAGTTGGT AATGGCAACG  560
TTTGTTGACT ACCGTATTGT AACTTTTGTC AAATTGTCAT AAATACGTGT CAAACTCTGG TAAAAAATTA  630
GTCTGCTACA TCTGTCTTTT ATTTATAAAA CACAGCTGTT AATCAGAATT TGGTTTATTA AATCAACAAC  700
CTGCACGAAA CTTGTGTGAG CATATTTTGT CTGTTTCTGG TTCATGACCT TCTTCCGCAT GATGGCCAAG  770
TGTAATGGCC ACTTGCAAGA GCGTTTCTTC AACGAGATAA GTCGAACAAA TATTTGTCCG TTACGACCAC  840
ATATAAAATC TCCCCATCTC TATATATAAT ACCAGCATTC ACCATCATGA ATACCTCAAA TCCCAATCTC  910
ACAAATACTT CAATAAAAAG ACCAAAAAAA ATTAAAGCAA AGAAAAGCCT TCTTGTGCAC AAAAAAAAAA  980
```

FIGURE 8
Page 1 of 4

```
GAAGCCTTCT AGGTTTTCAC GAC ATG AAG TTC ACT ACT CTA ATG GTC ATC ACA TTG   1036
                          MET Lys Phe Thr Thr Leu MET Val Ile Thr Leu

GTG ATA ATC GCC ATC TCG TCT CCT GTT CCA ATT AGA GCA ACC ACG GTT GAA AGT 1090
Val Ile Ile Ala Ile Ser Ser Pro Val Pro Ile Arg Ala Thr Thr Val Glu Ser

TTC GGA GAA GTG GCA CAA TCG TGT GTT GTG ACA GAA CTC GCC CCA TGC TTA CCA 1144
Phe Gly Glu Val Ala Gln Ser Cys Val Val Thr Glu Leu Ala Pro Cys Leu Pro

GCA ATG ACC ACG GCA GGA GAC CCG ACT ACA GAA TGC TGC GAC AAA CTG GTA GAG 1198
Ala MET Thr Thr Ala Gly Asp Pro Thr Thr Glu Cys Cys Asp Lys Leu Val Glu

CAG AAA CCA TGT CTT TGT GGT TAT ATT CGA AAC CCA GCC TAT AGT ATG TAT GTT 1252
Gln Lys Pro Cys Leu Cys Gly Tyr Ile Arg Asn Pro Ala Tyr Ser MET Tyr Val

ACT TCT CCA AAC GGT CGC AAA GTC TTA GAT TTT TGT AAG GTT CCT TTT CCT AGT 1306
Thr Ser Pro Asn Gly Arg Lys Val Leu Asp Phe Cys Lys Val Pro Phe Pro Ser

TGT TAAATCTCTC AAGACATTGC TAAGAAAAAT ATTATTAAAA ATAAAAGAAT CAAACTAGAT   1369
Cys

CTGATGTAAC AATGAATCAT CATGTTATGG TTGAAGCTTA TATGCTGAAG TGTTTGATTT TATATATGTG 1439

TGTGTGTGTG TCCTGCTCAA TTTTTGAAAC ACACACGTTT CTCCTGATTT GGATTTAAAT TATATTTTGA 1509

GTTAAAAAAA AGAAAAAGAT GGAATGCTAT TTATACAAGT TGATGAAAAA GTGGAAGTAC AATTTAGATA 1579
```

FIGURE 8
Page 2 of 4

```
TCTCCTACAC TTAAAGAATG AAACAATAAT AGACTTACGA AACAAATGAA AAATACATAA ATTGTCGACA  1649
ATCAACGTCC GATGACGAGT TTATTATTAA AAATTTGTGT GAAGGACTAG CAGTTCAACC AAATGATATT  1719
GAACATATAC ATCAACAAAT ATGATAATCA TAAAAGAGAG AATGGGGGGG GGGTGTCGTT TACCAGAAAC  1789
CTCTTTTTCT CAGCTCGCTA AAACCCTACC ACTAGAGACC TAGCTCTGAC CGTCGGCTCA TCGGTGCCGG  1859
AGGTGTAACC TTTCTTTCCC ATGACCCGAA ACCTCTCTTT CCCAACTCAC GAAAACCCTA CAATCAAAAA  1929
CCTAGCTCCG ACCGTCGGCT CATCGGTGCC GAAGGTGTAA CCTTTCTCTC CCATCATAGT TTCTCGTAAA  1999
TGAAAGCTAA TTGGGCAATC GATTTTTTAA TGTTTAAACC ATGCCAAGCC ATTTCTTATA GGACAATTGT  2069
CAATAATAGC ATCTTTTGAG TTTTGTCTCA AAAGTGACAC TAGAAGAAAA AAGTCACAAA AATGACATTC  2139
ATTAAAAAGT AAAATATCCC TAATACCTTT GGTTTAAATT AAATAAGTAA ACAAAAATAA ATAAAAACAA  2209
ATAAAATAAA AATAAAAAAT GAAAAAAAGA AATTTTTTTA TAGTTTCAGA TTATATGTTT TCAGATTCGA  2279
AATTTTTTAA ATTCCCTTTT TTAAATTTTC TTTTTTGAAA TTTTTTTTTT TGAAATTTTT TGAAACTGTT  2349
TTTAAAATTT TTATTTTTAA TTTTTTAGTA TTTATTTTTT ATTTTATAAA ATTTTAAACG CTAATTCCAA  2419
AACTCCCCCC CCCCCCCCCC CCCCAATTCT CTCCTAGTCT TTTTCTCTTT CTTATATTTG GGCTTCTATC  2489
TTCTCTTTTT TTTTCAGGCC CAAAGTATCA TGTGTAACAA CCGGTGTTCA AAAACGCGCC CGCCTGGCCG  2559
```

FIGURE 8
Page 3 of 4

```
TTTACTCGCC CGATTAAATG ATGATCGGAA GGCTGCCATG GCGAGGCGGA GGTAATCAGT GGTTCTAGGC 2629
GCTGAAACTA GAAAACCTTC AAAAATCGAA ATTTTAAGAG CTAAATCGGT GTTTATCTCA TGAATCTATT 2699
ATATTTAGTT GAAACTCACA AGAATCGGTT GTAAAAACTA TGAAATCGTG CAAAAAAAAT GAAGAACAAA 2769
ATATTCTCAG ATCTGGAAAA CACAGAGAAG AGGTTGAAGA TGAGGGTAAA ATCGTATTTT GTCATTCATT 2839
AAACTAAAAT CAAAAAAAAA TGATGCAAAA TTCAATGATA ATAACTCGAA CTCGCAACCA TATGCATCTT 2909
TAGACTGCGA CACGGACCAC TAGACTAAGC AATTTTAATG TTTATTCATC ACAGACCTAA TATATGTCTA 2979
AAACTAGGCG CCGAGTACGC CCCGCTTAAT CCCGAGTTTT TGTTAGCTCG CTAGACCCAG GGTCACCGCC 3049
CGACTAACGA GTAGCGTAAT TCTGAACTGG GGTAACAACA TAGAGAACAT CGCCGACCCT TCCCTGCCGA 3119
TGATGCCGCC TCCGATGAAC TTCCTGTAAC GCCTTCAGTT TCCATTGATT TTCCCCTTTA ATCTGATCAG 3189
TTCCATGTTT TATCCAACTC ATCCCACTCC GTAGCATTTA ATCGATCTCA TCATTTACAT ACATAACCAG 3259
TAGGAGGTCT CATATAAATT TGAACGTTTC CAGCGATGAA CAGTGCCAAT CTCTGCGAAA TCCATTTCTC 3329
TAAGCTCAGG GCTGGCGGCT GCAGCCCGGG GATCCACTAG TTCTAGGCGG CCGCACCGCG GTGGAGCTCC 3399
AATTCGCCCT ATAGTGAGTC GTATTACGCG CGCTCACTGG C                                3440
```

FIGURE 8
Page 4 of 4

24/42

```
    XhoI
    |
  1 CTCGAGAGCTGAAGGATTTTTTGTTAGAGATTCAACGACAGATGGACCCTTCCTCCACTAGGCAACTGC  69
    2

 70 AAGAACCTAACAATGCAAATATCACTCCTCCTCAGCCTTCAAGGAGCGTTAATAGGACTGGAACAAGCG 138

                                  BglII
                                  |
139 GTCAAGTGAGTAAATTTTCCTTCCAAGATAGATCTCTATGGTTCGGTTCATGAAGTTTGTGGTTTAATT 207
                                  169

208 GTGTAGCAACAGGATAGTGCAAGTGAGAATAGAGTTCGACCTCATCTACCTACCCCGGAACCTCTGAAT 276
277 GTATCCCCATTGAAGAAGAAGAGGGCAAATCCTGCACCCAGAAGGATAAAGAAATTTTGGACGCCTGAA 345
346 GAAGTGGCAGTTCTGAGGGAAGGAGTAAAAGAGTATGTCTACTACTACTACTCTATAATCAAGTTTCAA 414
415 GAAGCTGAGCTTGGCTCTCACTTTATATGTTTGATGTTGTTGTGCAGGTATGGTAAATCATGGAAAGAG 483
484 ATAAAGAATGCAAACCCTGAAGTATTGGCAGAGAGGACTGAGGTGAGAGAGCATGTCACTTTTGTGTTA 552
553 CTCATCTGAATTATCTTATATGCGAATTGTAAGTGGTACTAAAAGGTTTGTAACTTTTGGTAGGTGGAT 621
622 TTGAAGGATAAATGGAGGAACTTGCTTCGGTAGCGGTAACAAGTTTTATATTGCTATGAAGTTTTTTTG 690
691 CCTGCGTGACGTATCAGCAGCTGTGGAGAAGATGGTATTAGAAAGGGTCTTTTCACATTTTGTGTTGTG 759
```

FIGURE 9
Page 1 of 6

```
                                                      BglII
                                                      |
760  ACAAATATTAATTCGGCCGGTATGGTTTGGTTAAGACTTGTTGAGAGACGTGTGGGGTTTTTTGATGTA  828

829  TAATTAGTCTGTGTTTAGAACGAAACAAGACTTGTTGCGTATGCTTTTTTTAACTTGAGGGGGTTTGTT  897

                                                     BglII
                                                     |
898  GTTGTTAGTTAGGAACTTGACTTTGTCTCTTTCTCTCAAGATCTGATTGGTAAGGTCTGGGTGGTAGTA  966
                                                     937

967  CTGTTTGGTTTAATTTGTTTTGACTATTGAGTCACTGTGGCCCATTGACTTTAAATTAGGCTGGTATAT 1035

1036 TTTTTGGTTTAAAACCGGTCTGAGATAGTGCAATTTCGATTCAGTCAATTTTAAATTCTTCAAGGTAAT 1104

1105 GGGCTGAATACTTGTATAGTTTTAAGACTTAACAGGCCTTAAAAGGCCCATGTTATCATAAAACGTCAT 1173

                     HindIII
                     |
1174 TGTTTAGAGTGCACCAAGCTTATAAAATGTAGCCAGGCCTTAAAAGACTTAACAGGCCTTAAAAGACTT 1242
                     1190

1243 AACATTCCTTAAAAGGCCCATGTTATCATAAAACGTCATCGTTTTGAGTGCACCAAGCTAAATGTAGCC 1311

1312 AGGCCTTAAAAGACTTAACAGGCCTTAAAAGGCCCATGTTATCATAAAACGCCGTCGTTTTGAGTGCAC 1380
```

FIGURE 9
Page 2 of 6

```
     HindIII
     |
1381 CAAGCTTATAAATGTAGCCAGCTACCTCGGGACATCACGCTCTTTGTACACTCCGCCATCTCTCTCTCT 1449
     1383

     XhoI   BglII                           SalI
     |      |                               |
1450 CTCGAGCAGATCTCTCTCGGGAATATCGACAATGTCGACCACTTTCTGCTCTTCCGTCTCCATGCAAGC 1518
     1451   1458                            1484

1519 CACTTCTCTGGTAATCTCATCTCCTTCTTGTGTTCCCAGATCGCTCTGATCATACTTTCTTTTAGATCA 1587

1588 TTTGCCTCTGATCTGTTGCTTGATGTTTGTTAACTCTCCACGCATGTTTGATTATGTTGAGAATTAGAA 1656

1657 AAAAAATGTTAGCTTTACGAATCTTTAGTGATCATTTCAATTGGATTTGCAATCTTGTGTGACATTTGA 1725

1726 GGCTTGTGTAGATTTCGATCTGTATTCATTTTGAATCACAGCTATAATAGTCATTTGAGTAGTAGTGTT 1794

1795 TTTAAATGAACATGTTTTGTTGTATTGATGGAACAAACAGGCAGCAACAACGAGGATTAGTTTCCAGAA 1863

1864 GCCAGCTTTGGTTTCAACGACTAATCTCTCCTTCAACCTCCGCCGTTCAATCCCCACTCGTTTCTCAAT 1932

1933 CTCCTGCGCGGTATGTTCTCATTCTCAGCATTTATTTCGAGCTTGCTTGTCATGGTACTCTCTCTAATT 2001

2002 GTCTATTTGGTTTATTAGGCCAAACCAGAGACGGTTGAGAAAGTGTCTAAGATAGTTAAGAAGCAGCTA 2070
```

FIGURE 9
Page 3 of 6

```
2071 TCACTCAAAGACGACCAAAAGGTCGTTGCGGAGACCAAGTTTGCTGATCTTGGAGCAGATTCTCTCGAC 2139

2140 ACTGTAAGTCATCAATCATTCTCTTATGTGAATAAAGAGAACTTGAAGAGTTTGTTTTTAACATATTAA 2208

                                                             EcoRV
                                                             |
2209 CTGAGTGTTTTGCATGCAGGTTGAGATAGTGATGGGTTTAGAGGAAGAGTTTGATATCGAAATGGCTGA 2277
                                                             2264

                                                               SstI
                                                               |
2278 AGAGAAAGCTCAGAAGATTGCTACTGTGGAGGAAGCTGCTGAACTCATTGAAGAGCTCGTTCAACTTAA 2346
                                                               2335

2347 GAAGTAATTTTAGTATTAAGAGCAGCCAAGGCTTTGTTGGGTTTGTTGTTTTCATAATCTTCCTGTCAT 2415

2416 TTTCTTTTTCTTTAATGTGTCAAGCGACTCTGTTGGTTTAAAGTAGTATCTGTTTGCCATGGATCTCTC 2484

             SalI                          HindIII
             |                             |
2485 TCTATTTGTCGACTGAAAACTTTTGGTTTACACATGAAAGCTTGTTCTTGTTCTTTCTTAAATCGAAAT 2553
             2493                          2523

2554 GCCAAATGCGAGATTAGGGAATCTTGTATTAACACATACATAAGTCAAAGAGTAGGCCCTAAGATGACA 2622

2623 ATTTATAAACAATCCTATTCACATTGTATATACAGGTTATGATTATTCCCAATCAGCGTCAAAGAATCC 2691
```

FIGURE 9
Page 4 of 6

2692 AGCATCTTTCATCTCTGAATAGTAGACATTCTCCAAGTTCACATCTTCCTCCTGCACCAAAAACCAGTA 2760

2761 CTAAATCATGAACATTGCAATAATCACATGCCTAGGCGAGAGTTTTGGTGATGTGGTGTTAGTGATAGT 2829

2830 GATACTGATGGTGCTAGAGCGGTTAAGAAGGATTAACCTGGAAGAAGTCTGCAAGGAAAGTAACATAGA 2898

2899 GAAGAGGAAGATAGGAGTGGTAACAAACACTTGTGATCCCATACAGCCTCCCAGCATTTTTCAAATGTT 2967

2968 ATTTCCTTACATAAAGAAACAAGAGAAGTCTGACTAGATGATATTTATATAGGATAAGTGTTTTACCAT 3036

3037 AAGCCAAAGTGAGCGCCGTTTGCAAGAGCTAACCAGACAGTACACGTTTGGCATATATCTCATCAACAT 3105

3106 GATCTGAAAAGTAACATATCACAGTTAATGAACACAATGGTTACCTTGAGAAGCAAATCAAGACCTATA 3174

3175 ACAAGCCCAGAGATGAGGAAAGTCCGTGTCAACGCTTCACCGCCATTCGCGTAGTTTCCTTGGAAGACA 3243

3244 AAGGCCACCAACCAAACTTACTTCCAGAAACAACACTCCAAATGTTGTCAACAAAGTCAATAGATTCCA 3312

3313 AACTACTTCGTTACAGGGTTGTATAGATAATATAATAGAATAGTGGGAAGATAGTATAAATAAAATAAA 3381

3382 TAAAAGATCCTATCGGTAAATAGTTTATAATATCGGGGGCGTATATAAAGTATAAAAGAAACTCTTCTC 3450

3451 CAATCCGACCGTTGAAAATCACTCTCAATCTCTGGCGTAACGACCGGATCGTTCGCGCGTAATTTTCGC 3519

3520 TGCTATAAATAGAAACTTTCCTCTTCTGTTTCTCGATCAAAATTTTTTTTTGGAAAAATTAAGTTTGAA 3588

FIGURE 9
Page 5 of 6

```
3589 TCTATCGTAGATGCTGTGACAAAAAAAAATTGTTTTATCGAAGATGAGAAACATGAGGCCTGTTCATGC 3657

                          BamHI
                          |
3658 AAGGAACCAGACCACGGATCCATCTTCGCCGATGATGACGTCTCCTCTGATGAATCGTCACGCACGGAC 3726
                          3674

        BamHI
        |
3727 AGGATCCAACGCTGGACCAGCATCTAACGCCAAGAAAGCACAGACGAAAGCAGCAGCTCAGAGACTCGC 3795
        3729

3796 GGCTGTGATGTCGAACCAAACAGGCGACGATGAAGACAGTGATGATGACCTTTCCTTTGACTACAACGC 3864

                                         BglII
                                         |
3865 TGTCGGAAGCATTGGTCTCGCTGCCGGAAGATCT 3898
                                         3894
```

FIGURE 9
Page 6 of 6

```
Lambda CGN1-2

NCG-186 Linear          LENGTH = 4325

                                          SduI
     XhoI                              NlaIV
     AvaI                                 HgiJII     HindIII
     |                                 |  |          |
  1 CTCGAGGCAGTCACTAACATGAAGTTTGACGAGGAGCCCAACTATGGGAAGCTTATTTCTCTTTTCGAT   69
     2                                    39         50
                                       36

                             XbaI                      SacI
                             |                         |
 70 ACTCTAATTGAGCCGTGCGCTCTATCTAGACCAATTAGAATTGATGGAGCTCTAAAGGTTGCTGGCTGT  138
                             95                        121

               NdeI                          SspI                      NdeI
               |                             |                         |
139 TTTCTTGTTCATATGATTAACTTCTAAACTTGTGTATAAATATTCTCTGAAAGTGCTTCTTTTGGCATA  207
               150                           180                       206

                                         Ksp632I
                                         |
208 TGTAGGTTGGGCAAAAACGAGGAAGATTGCTTCTCAATTTGGAAGAGGATGAACAGCCGAAGAAGAAAA  276
                                         245
```

FIGURE 10
Page 1 of 13

31/42

```
                                XhoII
                                |
277 TAAGAATAGGCAGTCCTGCTACTCAATGGATCTCAGTCTATAACGGTCGTCGTCCCATGAAACAGAGGT 345
                                305

                                                           MmeI   EcoRV
                                                           |      |
346 AAAACATTTTTTGCATATACACTTTGAAAGTTCCTCACTAACTGTGTAATCTTTTGGTAGATATCACTA 414
                                                           401    408

                               SduI
                            MstI
                          BclI HgiAI                      HaeI
                          | |  |                          |
415 CAATGTCGGAGAGACAA3GGCTGMNCANCATATACAAAAGGGAAATGAAGATGGCCTTTTGATTAGCTG 483
                          437  442                        469
                            439

                                SduI
                                HgiJII
                                |
484 TGTAGCATCAGCAGCTAATCTCTGGGCTCTCATCATGGATGCTGGAACTGGATTCACTTCTCAAGTTTA 552
                                512
```

FIGURE 10
Page 2 of 13

```
          Cfr10I
       BbvII
       |   |
553 TGAGTTGTCACCGGTCTTCCTACACAAGGTAATAATCAGTTGAAGCAATTAAGAATCAATTTGATTTGT  621
       560
          563

622 AGTAAACTAAGAAGAACTTACCTTATGTTTTCCCCGCAGGACTGGATTATGGAACAATGGGAAAAGAAC  690

                                               SacI
                                               |
691 TACTATATAAGCTCCATAGCTGGTTCAGATAACGGGAGCTCTTTAGTTGTTATGTCAAAAGGTTAGTGT  759
                                               731

                      BbvII
                      |
760 TTAGTGAATAATAAACTTATACCACAAAGTCTTCATTGACTTATTTATATACTTGTTGTGAATTGCTAG  828
                      782

829 GAACTACTTATTCTCAGCAGTCATACAAAGTGAGTGACTCATTTCCGTTCAAGTGGATAAATAAGAAAT  897

898 GGAAAGAAGATTTTCATGTAACCTCCATGACAACTGCTGGTAATCGTTGGGGTGTGGTAATGTCGAGGA  966

              BclI
              |
967 ACTCTGGCTTCTCTGATCAGGTAGGTTTTTGTCTCTTATTGTCTGGTGTTTTTATTTTCCCCTGATAGT 1035
              981
```

FIGURE 10
Page 3 of 13

```
1036 CTAATATGATAAACTCTGCGTTGTGAAAGGTGGTGGAGCTTGACTTTTTGTACCCAAGCGATGGGATAC 1104

1105 ATAGGAGGTGGGAGAATGGGTATAGAATAACATCAATGGCAGCAACTGCGGATCAAGCAGCTTTCATAT 1173

      Tth111II                                                         ScaI
      |                                                                |
1174 TAAGCATACCAAAGCGTAAGATGGTGGATGAAACTCAAGAGACTCTCCGCACCACCGCCTTTCCAAGTA 1242
      1175                                                             1242

                                           XhoII
                                           |
1243 CTCATGTCAAGGTTGGTTTCTTTAGCTTTGAACACAGATTTGGATCTTTTTGTTTTGTTTCCATATACT 1311
                                           1285

1312 TAGGACCTGAGAGCTTTTGGTTGATTTTTTTTTCAGGACAAATGGGCGAAGAATCTGTACATTGCATCA 1380

                                     AflII
                                     |
1381 ATATGCTATGGCAGGACAGTGTGCTGATACACACTTAAGCATCATGTGGAAAGCCAAAGACAATTGGAG 1449
                                     1415

1450 CGAGACTCAGGGTCGTCATAATACCAATCAAAGACGTAAAACCAGACGCAACCTCTTTGGTTGAATGTA 1518

                                                                       SspI
                                                                       |
1519 ATGAAAGGGATGTGTCTTGGTATGTATGTACGAATAACAAAAGAGAAGATGGAATTAGTAGTAGAAATA 1587
                                                                       1587
```

FIGURE 10
Page 4 of 13

```
                                                                 EcoRV
                                                                 |
1588 TTTGGGAGCTTTTTAAGCCCTTCAAGTGTGCTTTTTATCTTATTGATATCATCCATTTGCGTTGTTTAA 1656
                                                                 1635

             XbaI
             |
1657 TGCGTCTCTAGATATGTTCCTATATCTTTCTCAGTGTCTGATAAGTGAAATGTGAGAAAACCATACCAA 1725
             1664

              SspI
              |
1726 ACCAAAATATTCAAATCTTATTTTTAATAATGTTGAATCACTCGGAGTTGCCACCTTCTGTGCCAATTG 1794
              1734                                                    1789

                                                                    EcoRI
                                                                    |
1795 TGCTGAATCTATCACACTAGAAAAAAACATTTCTTCAAGGTAATGACTTGTGGACTATGTTCTGAATTC 1863
                                                                    1859

                                                Eco57I
                                                |
1864 TCATTAAGTTTTTATTTTCTGAAGTTTAAGTTTTTACCTTCTGTTTTGAAATATATCGTTCATAAGATG 1932
                                                1904
```

FIGURE 10
Page 5 of 13

```
                                            SphI
                                            NspI
                                            |
1933 TCACGCCAGGACATGAGCTACACATCGCACATAGCATGCAGATCAGGACGATTTGTCACTCACTTCAAA 2001
                                            1971

                                                        SphI
                                                  NdeI  NspI         PmaCI
               Tth111II                           [AvaIII]  SspI AflIII
               |                                  ||     |   |    |  |
2002 CACCTAAGAGCTTCTCTCTCACAGCGCACACACATATGCATGCAATATTTACACGTGATCGCCATGCAA 2070
               2015                                2037       2048 2053
                                                  2036    2044       2056

                                        SecI
                                        |
2071 ATCTCCATTCTCACCTATAAATTAGAGCCTCGGCTTCACTCTTTACTCAAACCAAAACTCATCACTACA 2139
                                        2099

                                          Ksp632I
                                          |
2140 GAACATACACAAATGGCGAACAAGCTCTTCCTCGTCTCGGCAACTCTCGCCTTGTTCTTCCTTCTCACC 2208
               METAlaAsnLysLeuPheLeuValSerAlaThrLeuAlaLeuPhePheLeuLeuThr
                                          2171
```

FIGURE 10
Page 6 of 13

```
                                    SalI
                                     HindII                                 NaeI
               AccI                   AccI                                Cfr10I
                |                     |||                                   | |
2209 AATGCCTCCGTCTACAGGACGGTTGTGGAAGTCGACGAAGATGATGCCACAAATCCAGCCGGCCCATTT 2277
     AsnAlaSerValTyrArgThrValValGluValAspGluAspAspAlaThrAsnProAlaGlyProPhe
               2220                   2241                                2267
                                       2242                                 2269
                                     2240

                                                                 Tth111II
                                                                 HindIII          NlaIV
                                                                  |                |
2278 AGGATTCCAAAATGTAGGAAGGAGTTTCAGCAAGCACAACACCTGAAAGCTTGCCAACAATGGCTCCAC 2346
     ArgIleProLysCysArgLysGluPheGlnGlnAlaGlnHisLeuLysAlaCysGlnGlnTrpLeuHis
                                                                 2325             2342

                Tth111II                NlaIV                                   BbvII
                 |                       |                                        |
2347 AAGCAGGCAATGCAGTCCGGTAGTGGTCCAAGCTGGACCCTCGATGGTGAGTTTGATTTTGAAGACGAC 2415
     LysGlnAlaMETGlnSerGlySerGlyProSerTrpThrLeuAspGlyGluPheAspPheGluAspAsp
                2363                    2384                                    2415

                   NlaIV
                     ApaI  GsuI HaeI  NspBII                                  SacI
                   | |      |    |     |                                       Ksp632
                                                                             |  |
2416 GTGGAGAACCAACAACAGGGCCCGCAGCAGAGGCCACCGCTGCTCCAGCAGTGCTGCAACGAGCTCCAC 2484
     ValGluAsnGlnGlnGlnGlyProGlnGlnArgProProLeuLeuGlnGlnCysCysAsnGluLeuHis
                     2438  2444 2449  2455                                   2481
                   2436                                                         2484
```

FIGURE 10

Page 7 of 13

```
2485 CAGGAAGAGCCACTTTGCGTTTGCCCAACCTTGAAAGGAGCATCCAAAGCCGTTAAACAACAGATTCGA 2553
     GlnGluGluProLeuCysValCysProThrLeuLysGlyAlaSerLysAlaValLysGlnGlnIleArg

2554 CAACAACAGGGACAACAAATGCAGGGACAGCAGATGCAGCAAGTGATTAGCCGTATCTACCAGACCGCT 2622
     GlnGlnGlnGlyGlnGlnMETGlnGlyGlnGlnMETGlnGlnValIleSerArgIleTyrGlnThrAla

                                                                     SecI
                                                                 BbvII
                                                                 |   |
2623 ACGCACTTACCTAGAGCTTGCAACATCAGGCAAGTTAGCATTTGCCCCTTCCAGAAGACCATGCCTGGG 2691
     ThrHisLeuProArgAlaCysAsnIleArgGlnValSerIleCysProPheGlnLysThrMETProGly
                                                                 2684
                                                                     2687

     NlaIV                                                 SecI
       HgiJII                           XhoI               DsaI
       ApaI                             AvaI          AccI
     | |                                |             |    |
2692 CCCGGCTTCTACTAGATTCCAAACGAATATCCTCGAGAGTGTGTATACCACGGTGATATGAGTGTGGTT 2760
     ProGlyPheTyr .
       2694                             2724          2736
     2692                                                  2740

                  HpaI
                  HindII
                  |
2761 GTTGATGTATGTTAACACTACATAGTCATGGTGTGTGTTCCATAAATAATGTACTAATGTAATAAGAAC 2829
                  2774
```

FIGURE 10
Page 8 of 13

```
                AccI
                |
2830 TACTCCGTAGACGGTAATAAAAGAGAAGTTTTTTTTTTACTCTTGCTACTTTCCTATAAAGTGATGAT 2898
                2838

                                                                 SpeI
                                      VspI                      ScaI
                                      |                         ||
2899 TAACAACAGATACACCAAAAAGAAAACAATTAATCTATATTCACAATGAAGCAGTACTAGTCTATTGAA 2967
                                      2929                      2954
                                                                 2955

        NspI
     AflIII
     |   |
2968 CATGTCAGATTTTCTTTTTCTAAATGTCTAATTAAGCCTTCAAGGCTAGTGATGATAAAAGATCATCCA 3036
     2968
         2972

         XhoII
           NlaIV                       MmeI
         BamHI                         BclI
         | |                           |
3037 ATGGGATCCAACAAAGACTCAAATCTGGTTTTGATCAGATACTTCAAAACTATTTTTGTATTCATTAAA 3105
         3041                          3069
           3043
```

FIGURE 10
Page 9 of 13

```
                                   BbvII                                             Tth111
                                   |                                                 |
3106 TTATGCAAGTGTTCTTTTATTTGGTGAAGACTCTTTAGAAGCAAAGAACGACAAGCAGTAATAAAAAAA 3174
                                   3139                                              3174

                                                                               VspI
                                                                               |
3175 ACAAAGTTCAGTTTTAAGATTTGTTATTGACTTATTGTCATTTGAAAAATATAGTATGATATTAATATA 3243
                                                                               3237

        Tth111II                                          VspI
        |                                                 |
3244 GTTTTATTTATATAATGCTTGTCTATTCAAGATTTGAGAACATTAATATGATACTGTCCACATATCCAA 3312

        3250                                              3287

                                    NdeI         Tth111II
                                    |            |
3313 TATATTAAGTTTCATTTCTGTTCAAACATATGATAAGATGGTCAAATGATTATGAGTTTTGTTATTTAC 3381
                                    3341         3352

                              Eco57I                              Eco57I
                              |                                   |
3382 CTGAAGAAAAGATAAGTGAGCTTCGAGTTTCTGAAGGGTACGTGATCTTCATTTCTTGGCTAAAAGCGA 3450
                              3404                                3434

3451 ATATGACATCACCTAGAGAAAGCCGATAATAGTAAACTCTGTTCTTGGTTTTTGGTTTAATCAAACCGA 3519
```

FIGURE 10
Page 10 of 13

```
                                            Tth111II
      Cfr10I                               NdeI
      |                                    ||
3520 ACCGGTAGCTGAGTGTCAAGTCAGCAAACATCGCAAACCATATGTCAATTCGTTAGATTCCCGGTTTAA 3588
      3521                                 3560
                                            3561

          Cfr10I
          |
3589 GTTGTAAACCGGTATTTCATTTGGTGAAAACCCTAGAAGCCAGCCANCCTTTTTAATCTAATTTTTGCA 3657
          3597

                                                                 NlaIV
                                                                 HindII
                                                                HgiCI
                                                                 BspHI
                                                                |||
3658 AACGAGAAGTCACCACACCTCTCCACTAAAACCCTGAACCTTACTGAGAGAAGCAGAGNCANNAAAGAA 3726
                                                                 3717
                                                                3716
                                                                  3718

            Eco31I         PmaCI                     Ksp632I
            |              |                         |
3727 CAAATAAAACCCGAAGATGAGACCACCACGTGCGGCGGGACGTTCAGGGGACGGGGAGGAAGAGAATGR 3795
            3740           3756                      3781

3796 CGGCGG5MNTTTGGTGGCGGCGGCGGACGTTTTGGTGGCGGCGGTGGACGTTTTGGTGGCGGCGGTGGA 3864
```

FIGURE 10
Page 11 of 13

```
                 EcoRV
                 |
3865 CCTTTGGTGGTGGATATCGTGACGAAGGACCTCCCAGTGAAGTCATTGGTTCGTTTACTCTTTTCTTAG 3933
                 3880

                                               HindIII
                                                  AflII
                                               |  |
3934 TCGAATCTTATTCTTGCTCTGCTCGTTGTTTTACCGATAAAGCTTAAGACTTTATTGATAAAGTTCTCA 4002
                                                  3977
                                               3974

4003 GCTTTGAATGTGAATGAACTGTTTCCTGCTTATTAGTGTTCCTTTGTTTTGAGTTGAATCACTGTCTTA 4071

4072 GCACTTTTGTTAGATTCATCTTTGTGTTTAAGTTAAAAGGTAGAAACTTTGTGACTTGTCTCCGTTATG 4140

        HpaI
        HindII                         Tth111II
        |                              |
4141 ACAAGGTTAACTTTGTTGGTTATAACAGAAGTTGCGACCTTTCTCCATGCTTGTGAGGGTGATGCTGTG 4209
        4149                           4179

                     XhoII
                     |
4210 GACCAAGCTCTCTCAGGCGAAGATCCCTTACTTCAATGCCCCAATCTACTTGGAAAACAAGACACAGAT 4278
                     4231
```

FIGURE 10
Page 12 of 13

```
                                          SalI
                                        PstI
                        HindIII               HindII
                  XhoII    BspMI               AccI EcoRI
                  |        | |              | |||    |
4279 TGGGAAAGTTGATGAGATCCAAGCTTGGGCTGCAGGTCGACGAATTC 4325
                  4294     4302               4316 4321
                        4300                   4317
                                        4313
                                          4315
```

FIGURE 10
Page 13 of 13

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/01746

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5): C12N 1/21, 15/29, 15/82; C07H 15/12

U.S. CL.: 435/172.3, 240,4, 252.3; 536/27

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
U.S.	435/172.3, 240.4, 252.3 ; 536/27 800/205, DIG.69

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

USPTO AUTOMATED PATENT SYSTEM: DIALOG FILES BIOTECH AND PATENTS.
SEE ATTACHMENT FOR SEARCH TERMS

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
	SEE ATTACHED PAGES	

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
24 June 1991	07 AUG 1991
International Searching Authority	**Signature of Authorized Officer**
RO/US	P.Rhodes P.Rhodes

Form PCT/ISA/210 (second sheet) (Rev.11-87)

III. DOCUMENTS CONSID D TO BE RELEVANT (CONTINUED FROM THE SEC) SHEET)	PCT/US91/01746	
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	World Soybean Research Conference III: Proceedings (Westview Press): Shibles (ed); Published 1985; Goodman et al; "Biotechnology and its impact on future improvements in soybean production and use"; pages 261-271. See pages 264-265.	1-22, 34-37
Y	Journal of Lipid Research; Volume 26; Issued 1985; Mattson et al; "Comparison of effects of dietary saturated, monounsaturated, and poly-unsaturated fatty acids on plasma and lipids and lipoprotein in man"; pages 194-202. See entire document.	1-22, 34-37
Y	EP A 0,255,377 (KRIDL et al) 03 February 1988. See entire document.	1-22, 34-37
Y	Trends in Biotechnology; Volume 5; Issued February 1987; Knauf; "The application of genetic engineering to oilseed crops"; pages 40-47. See entire document.	1-22, 34-37
Y	Trends in Biotechnology; Volume 7; Issued May 1989; Battey et al; "Genetic engineering for plant oils: potential and limitations"; pages 122-126. See entire document.	1-22, 34-37
Y	US, A, 4,446,235 (SEEBURG) 01 May 1984. See entire document.	1-22, 34-37
Y	US, A, 4,394,443 (WEISSMAN et al) 19 July 1983. See entire document.	1-22, 34-37
Y	Methods in Enzymology; Volume 71; Issued 1981; McKeon et al; "Stearoyl-acyl carrier protein desaturase from safflower seeds"; pages 275-281. See entire document.	1-22, 34-37
Y	Archives of Biochemistry and Biophysics; Volume 162; Issued 1974; Javorski et al; "Fat metabolism in higher plants, properties of a soluble stearyl-acyl carrier protein destaurase from maturing <u>Carthamus tinctorius</u>"; pages 158-165. See entire document.	1-22, 34-37
Y	The Journal of Biological Chemistry; Volume 257, Number 20; Issued 25 October 1982; McKeon et al; "Purification and characterization of the stear yl-acyl carrier protein desaturase and the acyl-acyl carrier protein thioesterase from maturing seeds of safflow r"; pages 12141-12147. See entire document.	1-22, 34-37

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET) PCT/US91/01746

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
X	Proceedings of the Flax Institute USA; Volume 41, Number 3; Issued 1971, Downey et al; "Genetic control of fatty acid composition in oilseed crops"; pages 1-3. See entire document.	23-33, 38-45
X Y	EP, A, 0323753 (WONG et al) 12 July 1989. See entire document.	23-29, 31-33 38-39, 41-45 30, 40
X Y	Journal of the American Oil Chemists Society; Volume 61, Number 1; Issued January 1984; Wilcox et al; "Genetic alteration of soybean oil composition by a chemical mutagen"; pages 97-100. See entire document.	23, 25-27, 29, 31, 33, 38-45 30, 40
X Y	Journal of the American Oil Chemists Society; Volume 59, Number 5; Issued May 1982; Wolf et al; "Effect of temperature on soybean seed constituents: oil, protein, moisture, fatty acids, amino acids and sugars"; pages [illegible]. See entire document.	23, 25-27, 29, 31, 33, 38-45 30, 40
Y	Lipids; Volume 4, Number 6; Issued 1969; Inkpen et al; "Desaturation of palmitate and stearate by cell-free fractions from soybean cotyledons"; pages 539-543. See entire document.	30, 40
Y	The Journal of Biological Chemistry; Volume 241; Issued 25 Apr. 1966; Nagai et al; "Enzymatic desaturation of stearyl acyl carrier protein"; pages 1925-1927. See entire document.	30, 40
X Y	The Journal of Heredity; Volume 80, Number 3; Issued March 1989; Moore et al; "The inheritance of high oleic acid in peanut"; pages 252-253. See entire document.	23, 25-27, 29 31, 33, 38-45 30, 40
X Y	Crop Science; Volume 24; Issued November-December 1984; Carver et al; "Developmental changes in acyl-composition of soybean seed selected for high oleic acid concentration"; pages 1016-1019. See entire document.	23, 25-27, 29 31, 33, 38-45 30, 40
X Y	Bodman et al., "Soybeans and Soybean Products: Processing of edible soybean oil" published 1951 by Interscience Publishers, Inc. (N.Y.), pages 649-725, see only pages 702-709.	31 and 33 32

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/53, 15/82, C11B 1/00 C12Q 1/68

A1

(11) International Publication Number: WO 93/11245

(43) International Publication Date: 10 June 1993 (10.06.93)

(21) International Application Number: PCT/US92/10284

(22) International Filing Date: 3 December 1992 (03.12.92)

(30) Priority data:
07/804,259 4 December 1991 (04.12.91) US

(60) Parent Application or Grant
(63) Related by Continuation
US 07/804,259 (CIP)
Filed on 4 December 1991 (04.12.91)

(71) Applicant *(for all designated States except US)*: E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : BROWSE, John [GB/US]; Route 1, Box 39A, Palouse, WA 99161 (US). GRAU, Luis, Perez [ES/US]; 204 Woodhill Court, Wilmington, DE 19809 (US). KINNEY, Anthony, J. [GB/US]; 609 Lore Avenue, Wilmington, DE 19809 (US). PIERCE, John, W., Jr. [US/US]; 302 Jackson Boulevard, Wilmington, DE 19803 (US). WIERZBICKI, Anna, M. [US/US]; 1312 Idlewood Road, Wilmington, DE 19805 (US). YADAV, Narendra, S. [IN/US]; 127 Jade Drive, Wilmington, DE 19810 (US).

(74) Agents: FLOYD, Linda, Axamethy et al.; E.I. du Pont de Nemours and Company, Legal/Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States: AU, BR, CA, JP, RU, UA, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: FATTY ACID DESATURASE GENES FROM PLANTS

(57) Abstract

The preparation and use of nucleic acid fragments encoding fatty acid desaturase enzymes are described. The invention permits alteration of plant lipid composition. Chimeric genes incorporating such nucleic acid fragments with suitable regulatory sequences may be used to create transgenic plants with altered levels of unsaturated fatty acids.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LI	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

TITLE

FATTY ACID DESATURASE GENES FROM PLANTS

FIELD OF THE INVENTION

The invention relates to the preparation and use of nucleic acid fragments encoding fatty acid desaturase enzymes to modify plant lipid composition.

BACKGROUND OF THE INVENTION

Plant lipids have a variety of industrial and nutritional uses and are central to plant membrane function and climatic adaptation. These lipids represent a vast array of chemical structures, and these structures determine the physiological and industrial properties of the lipid. Many of these structures result either directly or indirectly from metabolic processes that alter the degree of unsaturation of the lipid. Different metabolic regimes in different plants produce these altered lipids, and either domestication of exotic plant species or modification of agronomically adapted species is usually required to economically produce large amounts of the desired lipid.

Plant lipids find their major use as edible oils in the form of triacylglycerols. The specific performance and health attributes of edible oils are determined largely by their fatty acid composition. Most vegetable oils derived from commercial plant varieties are composed primarily of palmitic (16:0), stearic (18:0), oleic (18:1), linoleic (18:2) and linolenic (18:3) acids. Palmitic and stearic acids are, respectively, 16- and 18-carbon-long, saturated fatty acids. Oleic, linoleic, and linolenic acids are 18-carbon-long, unsaturated fatty acids containing one, two, and three double bonds, respectively. Oleic acid is referred to as a mono-unsaturated fatty acid, while linoleic and linolenic acids are referred to as poly-unsaturated fatty acids. The relative amounts of saturated and

unsaturated fatty acids, in commonly used, edible vegetable oils are summarized below (Table 1):

TABLE 1

Percentages of Saturated and Unsaturated Fatty Acids in the Oils of Selected Oil Crops

	Saturated	Mono-unsaturated	Poly-unsaturated
Canola	6%	58%	36%
Soybean	15%	24%	61%
Corn	13%	25%	62%
Peanut	18%	48%	34%
Safflower	9%	13%	78%
Sunflower	9%	41%	51%
Cotton	30%	19%	51%

Many recent research efforts have examined the role that saturated and unsaturated fatty acids play in reducing the risk of coronary heart disease. In the past, it was believed that mono-unsaturates, in contrast to saturates and poly-unsaturates, had no effect on serum cholesterol and coronary heart disease risk. Several recent human clinical studies suggest that diets high in mono-unsaturated fat and low in saturated fat may reduce the "bad" (low-density lipoprotein) cholesterol while maintaining the "good" (high-density lipoprotein) cholesterol (Mattson et al., Journal of Lipid Research (1985) 26:194-202).

A vegetable oil low in total saturates and high in mono-unsaturates would provide significant health benefits to consumers as well as economic benefits to oil processors. As an example, canola oil is considered a very healthy oil. However, in use, the high level of poly-unsaturated fatty acids in canola oil renders the oil unstable, easily oxidized, and susceptible to development of disagreeable odors and flavors (Gailliard, 1980, Vol. 4, pp. 85-116 In: Stumpf, P. K., Ed., The Biochemistry of Plants, Academic Press, New

York). The levels of poly-unsaturates may be reduced by hydrogenation, but the expense of this process and the concomitant production of nutritionally questionable trans isomers of the remaining unsaturated fatty acids reduces the overall desirability of the hydrogenated oil (Mensink et al., New England J. Medicine (1990) N323: 439-445). Similar problems exist with soybean and corn oils.

For specialized uses, high levels of poly-unsaturates can be desirable. Linoleate and linolenate are essential fatty acids in human diets, and an edible oil high in these fatty acids can be used for nutritional supplements, for example in baby foods. Linseed oil, derived from the Flax plant (Linum usitatissimum), contains over 50% linolenic acid and has widespread use in domestic and industrial coatings since the double bonds of the fatty acids react rapidly with oxygen to polymerize into a soft and flexible film. Although the oil content of flax is comparable to canola (around 40% dry weight of seed), high yields are only obtained in warm temperatures or subtropical climates. In the USA flax is highly susceptible to rust infection. It will be commercially useful if a crop such as soybean or canola could be genetically transformed by the appropriate desaturase gene(s) to synthesize oils with a high linolenic acid content.

Mutation-breeding programs have met with some success in altering the levels of poly-unsaturated fatty acid levels found in the edible oils of agronomic species. Examples of commercially grown varieties are high (85%) oleic sunflower and low (2%) linolenic flax (Knowles, (1980) pp. 35-38 In: Applewhite, T. H., Ed., World Conference on Biotechnology for the Fats and Oils Industry Proceedings, American Oil Chemists' Society). Similar commercial progress with the other plants shown

in Table 1 has been largely elusive due to the difficult nature of the procedure and the pleiotropic effects of the mutational regime on plant hardiness and yield potential.

The biosynthesis of the major plant lipids has been the focus of much research (Browse et al., Ann. Rev. Plant Physiol. Mol. Biol. (1991) 42:467-506). These studies show that, with the notable exception of the soluble stearoyl-acyl carrier protein desaturase, the controlling steps in the production of unsaturated fatty acids are largely catalyzed by membrane-associated fatty acid desaturases. Desaturation reactions occur in plastids and in the endoplasmic reticulum using a variety of substrates including galactolipids, sulfolipids, and phospholipids. Genetic and physiological analyses of Arabidopsis thaliana nuclear mutants defective in various fatty acid desaturation reactions indicates that most of these reactions are catalyzed by enzymes encoded at single genetic loci in the plant. The analyses show further that the different defects in fatty acid desaturation can have profound and different effects on the ultra-structural morphology, cold sensitivity, and photosynthetic capacity of the plants (Ohlrogge, et al., Biochim. Biophys. Acta (1991) 1082:1-26). However, biochemical characterization of the desaturase reactions has been meager. The instability of the enzymes and the intractability of their proper assay has largely limited researchers to investigations of enzyme activities in crude membrane preparations. These investigations have, however, demonstrated the role of delta-12 desaturase and delta-15 desaturase activities in the production of linoleate and linolenate from 2-oleoyl-phosphatidyl-choline and 2-linoleoyl-phosphatidylcholine, respectively (Wang et al., Plant Physiol. Biochem.

(1988) 26:777-792). Thus, modification of the activities of these enzymes represents an attractive target for altering the levels of lipid unsaturation by genetic engineering.

Genes from plants for stearoyl-acyl carrier protein desaturase, the only soluble fatty acid desaturase known, have been described (Thompson, et al., Proc. Natl. Acad. Sci. U.S.A. (1991) 88:2578-2582; Shanklin et al., Proc. Natl. Acad. Sci. USA (1991) 88:2510-2514). Stearoyl-coenzyme-A desaturase genes from yeast, rat, and mice have also been described (Stukey, et al., J. Biol. Chem.(1990) 265:20144-20149; Thiede, et al., J. Biol. Chem. (1986) 261:13230-13235; Kaestner, et al., J. Biol. Chem. (1989) 264:14755-1476). No evidence exists in the public art that describes the isolation of fatty acid desaturases other than stearoyl-ACP desaturases from higher plants or their corresponding genes. A fatty acid desaturase gene from the cyanobacterium, Synechocystis PCC 6803, has also been described (Wada, et al., Nature (1990) 347:200-203). This gene encodes a fatty acid desaturase, designated des A, that catalyzes the conversion of oleic acid at the 1 position of galactolipids to linoleic acid. However, these genes have not proven useful for isolating plant fatty acid desaturases other than stearoyl-ACP desaturase via sequence-dependent protocols, and the present art does not indicate how to obtain plant fatty acid desaturases other than stearoyl-ACP desaturases or how to obtain fatty acid desaturase-related enzymes. Thus, the present art does not teach how to obtain glycerolipid desaturases from plants. Furthermore, there is no evidence that a method to control the nature and levels of unsaturated fatty acids in plants using nucleic acids encoding fatty acid desaturases other than stearoyl-ACP desaturase is known in the art.

The biosynthesis of the minor plant lipids has been less well studied. While hundreds of different fatty acids have been found, many from the plant kingdom, only a tiny fraction of all plants have been surveyed for their lipid content (Gunstone, et al., Eds., (1986) The Lipids Handbook, Chapman and Hall Ltd., Cambridge). Accordingly, little is known about the biosynthesis of these unusual fatty acids and fatty acid derivatives. Interesting chemical features found in such fatty acids include, for example, allenic and conjugated double bonds, acetylenic bonds, trans double bonds, multiple double bonds, and single double bonds in a wide number of positions and configurations along the fatty acid chain. Similarly, many of the structural modifications found in unusual lipids (e.g., hydroxylation, epoxidation, cyclization, etc.) are probably produced via further metabolism following chemical activation of the fatty acid by desaturation or they involve a chemical reaction that is mechanistically similar to desaturation. For example, evidence for the mechanism of hydroxylation of fatty acids being part of a general mechanism of enzyme-catalyzed desaturation in eukaryotes has been obtained by substituting a sulfur atom in the place of carbon at the delta-9 position of stearic acid. When incubated with yeast cell extracts the thiostearate was converted to a 9-sulfoxide (Buist et al. (1987) Tetrahedron Letters 28:857-860). This sulfoxidation was specific for sulfur at the delta-9 position and did not occur in a yeast delta-9-desaturase deficient mutant (Buist & Marecak (1991) Tetrahedron Letters 32:891-894). The 9-sulfoxide is the sulfur analogue of 9-hydroxyoctadecastearate, the proposed intermediate of stearate desaturation. Thus fatty-acid desaturase cDNAs may serve as useful probes for cDNAs encoding fatty-acid hydroxylases and other cDNAs which encode enzymes with

reaction mechanisms similar to fatty-acid desaturation. Many of these fatty acids and derivatives having such features within their structure could prove commercially useful if an agronomically viable species could be induced to synthesize them by introduction of a gene encoding the appropriate desaturase.

SUMMARY OF THE INVENTION

Applicants have discovered a means to control the nature and levels of unsaturated fatty acids in plants. Nucleic acid fragments from glycerolipid desaturase cDNAs or genes are used to create chimeric genes. The chimeric genes may be used to transform various plants to modify the fatty acid composition of the plant or the oil produced by the plant. More specifically, one embodiment of the invention is an isolated nucleic acid fragment comprising a nucleotide sequence encoding a plant delta-15 fatty acid desaturase or a fatty acid desaturase-related enzyme with an amino acid identity of 50%, 65%, 90% or greater to the polypeptide encoded by SEQ ID NOS:1, 4, 6, 8, 10, 12, 14, or 16. The isolated fragment in these embodiments is isolated froma plant selected from the group consisting of soybean, oilseed Brassica species, Arabidopsis thaliana and corn.

Another embodiment of this invention involves the use of these nucleic acid fragments in sequence-dependent protocols. Examples include use of the fragments as hybridization probes to isolate other glycerolipid desaturase cDNAs or genes. A related embodiment involves using the disclosed sequences for amplification of DNA fragments encoding other glycerolipid desaturases.

Another aspect of this invention involves chimeric genes capable of causing altered levels of the linolenic acid in a transformed plant cell, the gene comprising nucleic acid fragments encoding encoding a plant

delta-15 fatty acid desaturase or a fatty acid desaturase-related enzyme with an amino acid identity of 50%, 65%, 90% or greater to the polypeptide encoded by SEQ ID NOS:1, 4, 6, 8, 10, 12, 14, or 16 operably linked in suitable orientation to suitable regulatory sequences. Preferred are those chimeric genes which incorporate nucleic acid fragments encoding delta-15 fatty acid desaturase cDNAs or genes. Plants and oil from seeds of plants containing the chimeric genes described are also claimed.

Yet another embodiment of the invention involves a method of producing seed oil containing altered levels of linolenic (18:3) acid comprising: (a) transforming a plant cell with a chimeric gene described above; (b) growing fertile plants from the transformed plant cells of step (a); (c) screening progeny seeds from the fertile plants of step (b) for the desired levels of linolenic (18:3) acid, and (d) processing the progeny seed of step (c) to obtain seed oil containing altered levels of the unsaturated fatty acids. Preferred plant cells and oils are derived from soybean, rapeseed, sunflower, cotton, cocoa, peanut, safflower, coconut, flax, oil palm, and corn. Preferred methods of transforming such plant cells would include the use of Ti and Ri plasmids of Agrobacterium, electroporation, and high-velocity ballistic bombardment.

The invention also is embodied in a method of breeding plant species to obtain altered levels of polyunsaturated fatty acids, specifically linolenic (18:3) acid in seed oil of oil-producing plants. This method involves (a) making a cross between two varieties of an oilseed plant differing in the linolenic acid trait; (b) making a Southern blot of restriction enzyme digested genomic DNA isolated from several progeny plants resulting from the cross of step (a); and (c)

hybridizing the Southern blot with the radiolabeled nucleic acid fragments encoding the claimed glycerolipid desaturases.

The invention is also embodied in a method of RFLP mapping that uses the isolated *Arabidopsis thaliana* delta-15 desaturase sequences described herein.

The invention is also embodied in plants capable of producing altered levels of glycerolipid desaturase by virtue of containing the chimeric genes described herein. Further, the invention is embodied by seed oil obtained from such plants.

The invention is also embodied in a method of RFLP mapping ina genomic RFLP marker comprising(a) making a cross between two varieties of plants; (b)making a Southern blot of restriction enzyme digested genomic DNA isolated from several progeny plants resulting from the cross of step (a); and (c) hybridizing the Southern blot with a radiolabelled nucleic acid fragments of the claimed fragments.

The invention is also embodied in a method to isolate nucleic acid fragments encoding fatty acid desaturases and fatty acid desaturase-related enzymes, comprising (a)comparing SEQ ID NOS:2, 5, 7, 9, 11, 13, 15 and 17 with other fatty acid desaturase polypeptide sequences; (b) identifying the conserved sequence(s) of 4 or more amino acids obtained in step a; (c) making region-specific nucleotide probe(s) or oligomer(s) based on the conserved sequences identified in step b; and d) using the nucleotide probe(s) or oligomers(s) of step c to isolate sequences encoding fatty acid desaturases and fatty-acid desaturase-related enzymes by sequence-dependent protocols. The product of the method of isolation method described is also part of the invention.

BRIEF DESCRIPTION OF THE SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the Sequence Descriptions which form a part of this application. The Sequence Descriptions contain the one letter code for nucleotide sequence characters and the three letter code for amino acids in conformity with the IUPAC-IUB standard described in Nucleic Acids Research 13:3021-3030 (19085) and 37 C.F.R. 1.822 which are incorporated herein by reference.

SEQ ID NO:1 shows the complete 5' to 3' nucleotide sequence of 1350 base pairs of the *Arabidopsis* cDNA which encodes delta-15 desaturase in plasmid pCF3. Nucleotides 46 to 48 are the putative initiation codon of the open reading frame (nucleotides 46 to 1206). Nucleotides 1204 to 1206 are the termination codon. Nucleotides 1 to 45 and 1207 to 1350 are the 5' and 3' untranslated nucleotides, respectively. The 386 amino acid protein sequence in SEQ ID NO:1 is that deduced from the open reading frame.

SEQ ID NO:2 is the deduced peptide of the open-reading frame of SEQ ID NO:1.

SEQ ID NO:3 is a partial nucleotide sequence of the *Arabidopsis* genomic DNA insert in plasmid pF1 which shows the genomic sequence in the region of the *Arabidopsis* genome that encodes delta-15 desaturase. Nucleotides 68-255 are identical to nucleotides 1-188 of SEQ ID NO:1. Nucleotides 47 to 49 and 56 to 58 are termination codons in the same reading frame as the open reading frame in SEQ ID NO:1.

SEQ ID NO:4 shows the 5' to 3' nucleotide sequence of the insert in plasmid pACF2-2 of 1525 base pairs of the *Arabidopsis thaliana* cDNA that encodes a plastid delta-15 fatty acid desaturase. Nucleotides 10-12 and nucleotides 1348 to 1350 are, respectively, the putative

initiation codon and the termination codon of the open reading frame (nucleotides 10 to 1350). Nucleotides 1 to 9 and 1351 to 1525 are, respectively, the 5' and 3' untranslated nucleotides.

SEQ ID NO:5 is the deduced peptide of the open reading frame of SEQ ID NO:4.

SEQ ID NO:6 shows the complete 5' to 3' nucleotide sequence of 1336 base pairs of the Brassica napus seed cDNA, found in plasmid pBNSF3-2, which encodes a microsomal delta-15 glycerolipid desaturase. Nucleotides 79 to 82 are the putative initiation codon of the open reading frame (nucleotides 79 to 1212). Nucleotides 1210 to 1212 are the termination codon. Nucleotides 1 to 78 and 1213 to 1336 are the 5' and 3' unstranslated nucleotides respectively.

SEQ ID NO:7 is the deduced peptide of the open reading frame of SEQ ID NO:6.

SEQ ID NO:8 is the complete 5' to 3' nucleotide sequence of 1416 base pairs of the Brassica napus seed cDNA found in plasmid pBNSFd-2 which encodes a plastid delta-15 glycerolipid desaturase. Nucleotides 1 to 1215 correspond to a continuous open reading frame of 404 amino acids. Nucleotides 1213 to 1215 are the termination codon. Nucleotides 1215 to 1416 are the 3' untranslated nucleotides.

SEQ ID NO:9 is the deduced peptide of the open reading frame of SEQ ID NO:8.

SEQ ID NO:10 is the complete nucleotide sequence of the soybean (glycine max) microsomal delta-15 desaturase cDNA, found in plasmid pXF1, which the 2184 nucleotides of this sequence contain·both the coding sequence and the 5' and 3' non-translated regions of the cDNA. Nucleotides 855 to 857 are the putative initiation codon of the open reading frame (nucleotides 855 to 2000). Nucleotides 1995 to 1997 are the termination codon.

Nucleotides 1 to 854 and 1998 to 2184 are the 5' and 3' unstranslated nucl otides respectively. The 380 amino acid protein sequence in SEQ ID NO:7 is that deduced from the open reading frame.

SEQ ID NO:11 is the deduced peptide of the open reading frame in SEQ ID NO:10.

SEQ ID NO:12 is the complete 5' to 3' nucleotide sequence of 1676 base pairs of the soybean (<u>Glycine max</u>) seed cDNA found in plasmid pSFD-118bwp which encodes a soybean plastid delta-15 desaturase. Nucleotides 169 to 1530 correspond to a continuous open reading frame of 453 amino acids. Nucleotides 169 to 171 are the putative initiation codon of the open reading frame. Nucleotides 1528 to 1530 are the termination codon. Nucleotides 1531 to 1676 are the 3' untranslated nucleotides. Nucleotides 169 to 382 encode the putative plastid transit peptide, based on comparison of the deduced peptide with the soybean microsomal delta-15 peptide.

SEQ ID NO:13 is the deduced peptide of the open reading frame in SEQ ID NO:12.

SEQ ID NO:14 is the complete nucleotide sequence of a 396 bp polymerase chain reaction product derived from corn seed mRNA that is found in the insert of plasmid pPCR20. Nucleotides 1 to 31 and 364 to 396 correspond to the amplification primers described in SEQ ID NO:18 and SEQ ID NO:19, respectively. Nucleotides 31 to 363 encode an internal region of a corn seed delta-15 desaturase that is 61.9% identical to the region between amino acids 137 and 249 of the Brassica napus delta-15 desaturase peptide sequence shown in SEQ ID NO:7.

SEQ ID NO:15 is the deduced amino acid sequence of SEQ ID NO:14.

SEQ ID NO:16 shows the partial composite 5' to 3' nucleotide sequence of 472 bp derived from the inserts

in plasmids pFadx-2 and pYacp7 for Arabidopsis thaliana cDNA that encodes a plastid delta-15 fatty acid desaturase. Nucleotides 2-4 and nucleotides 468 to 470 are, respectively, the first and the last codons in the open reading frame.

SEQ ID NO:17 is deduced partial peptide sequence of the open reading frame in SEQ ID NO:16.

SEQ ID NO:18 One hundred and twenty eight fold degenerate sense 31-mer PCR primer. Nucleotides 1 to 8 correspond to the Bam H1 restriction enzyme recognition sequence. Nucleotides 9 to 137 correspond to amino acid residues 130 to 137 of SEQ ID NO:6 with a deoxyinosine base at nucleotide 11.

SEQ ID NO:19 Two thousand and forty eight-fold degenerate antisense 35-mer PCR primer. Nucleotides 1 to 8 correspond to the Bam H1 restriction enzyme recognition sequence. Nucleotides 9 to 35 correspond to amino acid residues 249 to 256 of SEQ ID NO:6 with a deoxyinosine base at nucleotide 15.

SEQ ID NO:20 Sixteen-fold degenerate sense 36-mers made to amino acid residues 97-108 in SEQ ID NO:2.

SEQ ID NO:21 Sixteen-fold degenerate sense 36-mers made to amino acid residues 97-108 in SEQ ID NO:2.

SEQ ID NO:22 Seventy two-fold degenerate sense 18-mers made to amino acid residues 100-105 in SEQ ID NO:2.

SEQ ID NO:23 Seventy two-fold degenerate sense 18-mers made to amino acid residues 100-105 in SEQ ID NO:2.

SEQ ID NO:24 Seventy two-fold degenerate antisense 18-mers made to amino acid residues 299-304 in SEQ ID NO:2.

SEQ ID NO:25 Seventy two-fold degenerate antisense 18-mers made to amino acid residues 299-304 in SEQ ID NO:2.

SEQ ID NO:26 S venty two-fold degenerate antisense 18-mers made to amino acid residues 304-309 in SEQ ID NO:2.

SEQ ID NO:27 Seventy two-fold degenerate antisense 18-mers made to amino acid residues 304-309 in SEQ ID NO:2.

SEQ ID NO:28 Sixteen-fold degenerate sense 36-mers made to amino acid residues 97-108 in SEQ ID NO:2.

SEQ ID NO:29 Sixteen-fold degenerate sense 36-mers made to amino acid residues 97-108 in SEQ ID NO:2.

SEQ ID NO:30 Sixty four-fold degenerate antisense 38-mers made to amino acid residues 299-311 in SEQ ID NO:2.

SEQ ID NO:31 Sixty four-fold degenerate antisense 38-mers made to amino acid residues 299-311 in SEQ ID NO:2.

SEQ ID NO:32 A 135-mer made as an antisense strand to amino acid residues 97-141 in SEQ ID NO:2.

DETAILED DESCRIPTION OF THE INVENTION

Applicants have isolated nucleic acid fragments that encode plant fatty acid desaturases and that are useful in modifying fatty acid composition in oil-producing species by transformation.

Thus, transfer of the nucleic acid fragments of the invention or a part thereof that encodes a functional enzyme, along with suitable regulatory sequences that direct the transciption of their mRNA, into a living cell will result in the production or over-production of plant fatty acid desaturases and will result in increased levels of unsaturated fatty acids in cellular lipids, including triacylglycerols.

Transfer of the nucleic acid fragments of the invention or a part thereof, along with suitable regulatory sequences that direct the transciption of their antisense RNA, into plants will result in the

inhibition of expression of the endogenous fatty acid desaturase that is substantially homologous with the transferred nucleic acid fragment and will result in decreased levels of unsaturated fatty acids in cellular lipids, including triacylglycerols.

Transfer of the nucleic acid fragments of the invention or a part thereof, along with suitable regulatory sequences that direct the transciption of their mRNA, into plants may result in inhibition by cosuppression of the expression of the endogenous fatty acid desaturase gene that is substantially homologous with the transferred nucleic acid fragment and may result in decreased levels of unsaturated fatty acids in cellular lipids, including triacylglycerols.

The nucleic acid fragments of the invention can also be used as restriction fragment length polymorphism (RFLP) markers in Arabidopsis genetic mapping and plant breeding programs.

The nucleic acid fragments of the invention or oligomers derived therefrom can also be used to isolate other related glycerolipid desaturase genes using DNA, RNA, or a library of cloned nucleotide sequences from the same or different species by well known sequence-dependent protocols, including, for example, methods of nucleic acid hybridization and amplification by the polymerase chain reaction.

Definitions

In the context of this disclosure, a number of terms shall be used. The term "fatty acid desaturase" used herein refers to an enzyme which catalyzes the breakage of a carbon-hydrogen bond and the introduction of a carbon-carbon double bond into a fatty acid molecule. The fatty acid may be free or esterified to another molecule including, but not limited to, acyl-carrier protein, coenzyme A, sterols and the glycerol

moiety of glycerolipids. The term "glycerolipid desaturases" used herein ref rs to a subset of the fatty acid desaturases that act on fatty acyl moieties esterified to a glycerol backbone. "Delta-12 desaturase" refers to a fatty acid desaturase that catalyzes the formation of a double bond between carbon positions 6 and 7 (numbered from the methyl end), (i.e., those that correspond to carbon positions 12 and 13 (numbered from the carbonyl carbon) of an 18 carbon-long fatty acyl chain or carbon positions 10 and 11 (numbered from the carbonyl carbon) of a 16 carbon-long fatty acyl chain). "Delta-15 desaturase" refers to a fatty acid desaturase that catalyzes the formation of a double bond between carbon positions 3 and 4 (numbered from the methyl end), (i.e., those that correspond to carbon positions 15 and 16 (numbered from the carbonyl carbon) of an 18 carbon-long fatty acyl chain and carbon positions 13 and 14 (numbered from the carbonyl carbon) of a 16 carbon-long fatty acyl chain). Examples of fatty acid desaturases include, but are not limited to, the microsomal delta-12 and delta-15 desaturases that act on phosphatidylcholine lipid substrates; the chloroplastic delta-12 and delta-15 desaturases that act on phosphatidyl glycerol and galactolipids; and other desaturases that act on such fatty acid substrates such as phospholipids, galactolipids, and sulfolipids. "Microsomal desaturase" refers to the cytoplasmic location of the enzyme, while "chloroplast desaturase" and "plastid desaturase" refer to the plastid location of the enzyme. These fatty acid desaturases may be found in a variety of organisms including, but not limited to, higher plants, diatoms, and various eukaryotic and prokaryotic microorganisms such as fungi and photosynthetic bacteria and algae. The term "homologous fatty acid desaturases" refers to fatty acid

desaturases that catalyze the same desaturation on the same lipid substrate. Thus, microsomal delta-15 desaturases, even from different plant species, are homologous fatty acid desaturases. The term "heterologous fatty acid desaturases" refers to fatty acid desaturases that catalyze desaturations at different positions and/or on different lipid substrates. Thus, for example, microsomal delta-12 and delta-15 desaturases, which act on phosphatidylcholine lipids, are heterologous fatty acid desaturases, even when from the same plant. Similarly, microsomal delta-15 desaturase, which acts on phosphatidylcholine lipids, and chloroplast delta-15 desaturase, which acts on galactolipids, are heterologous fatty acid desaturases, even when from the same plant. It should be noted that these fatty acid desaturases have never been isolated and characterized as proteins. Accordingly the terms such as "delta-12 desaturase" and "delta-15 desaturase" are used as a convenience to describe the proteins encoded by nucleic acid fragments that have been isolated based on the phenotypic effects caused by their disruption. The term "fatty acid desaturase-related enzyme" refers to enzymes whose catalytic product may not be a carbon-carbon double bond but whose mechanism of action is similar to that of a fatty acid desaturase (that is, catalysis of the displacement of a carbon-hydrogen bond of a fatty acid chain to form a fatty-hydroxyacyl intermediate or end-product). This term is different from "related fatty acid desaturases", which refers to structural similarities between fatty acid desaturases.

The term "nucleic acid" refers to a large molecule which can be single-stranded or double-stranded, composed of monomers (nucleotides) containing a sugar, a phosphate and either a purine or pyrimidine. A "nucleic

acid fragment" is a fraction of a given nucleic acid molecule. In higher plants, deoxyribonucleic acid (DNA) is the genetic material while ribonucleic acid (RNA) is involved in the transfer of the information in DNA into proteins. A "genome" is the entire body of genetic material contained in each cell of an organism. The term "nucleotide sequence" refers to the sequence of DNA or RNA polymers, which can be single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers. The term "oligomer" refers to short nucleotide sequences, usually up to 150 bases long. "Region-specific nucleotide probes" refers to isolated nucleic acid fragments derived from a cDNA or gene using a knowledge of the amino acid regions conserved between different fatty-acid desaturases which may be used to isolate cDNAS or genes for other fatty-acid desaturases or fatty acid desaturase-related enzymes using sequence dependent protocols. As used herein, the term "homologous to" refers to the relatedness between the nucleotide sequence of two nucleic acid molecules or between the amino acid sequences of two protein molecules. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) Nucleic Acid Hybridisation, IRL Press, Oxford, U.K.); or by the comparison of sequence similarity between two nucleic acids or proteins, such as by the method of Needleman et al. (J. Mol. Biol. (1970) 48:443-453). As used herein, "substantially homologous" refers to nucleotide sequences that have more than 90% overall identity at the nucleotide level with the coding region of the claimed sequence, such as genes and pseudo-genes corresponding to the coding

regions. The nucleic acid fragments described herein include molecules which comprise possible variations, both man-made and natural, such as but not limited to (a) those that involve base changes that do not cause a change in an encoded amino acid, or (b) which involve base changes that alter an amino acid but do not affect the functional properties of the protein encoded by the DNA sequence, (c) those derived from deletions, rearrangements, amplifications, random or controlled mutagenesis of the nucleic acid fragment, and (d) even occasional nucleotide sequencing errors.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding) and following (3' non-coding) the coding region. "Fatty acid desaturase gene" refers to a nucleic acid fragment that expresses a protein with fatty acid desaturase activity. "Native" gene refers to an isolated gene with its own regulatory sequences as found in nature. "Chimeric gene" refers to a gene that comprises heterogeneous regulatory and coding sequences not found in nature. "Endogenous" gene refers to the native gene normally found in its natural location in the genome and is not isolated. A "foreign" gene refers to a gene not normally found in the host organism but that is instead introduced by gene transfer. "Pseudo-gene" refers to a genomic nucleotide sequence that does not encode a functional enzyme.

"Coding sequence" refers to a DNA sequence that codes for a specific protein and excludes the non-coding sequences. It may constitute an "uninterrupted coding sequence", i.e., lacking an intron or it may include one or more introns bounded by appropriate splice junctions. An "intron" is a nucleotide sequence that is transcribed in the primary transcript but that is removed through cleavage and re-ligation of the RNA within the cell to

create the mature mRNA that can be translated into a protein.

"Initiation codon" and "termination codon" refer to a unit of three adjacent nucleotides in a coding sequence that specifies initiation and chain termination respectively, of protein synthesis (mRNA translation). "Open reading frame" refers to the coding sequence uninterrupted by introns between initiation and termination codons that encodes an amino acid sequence.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene by interfering with the processing, transport and/or translation of its primary transcript or mRNA. The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. In addition, as used herein, antisense RNA may contain regions of ribozyme sequences that increase the efficacy of antisense RNA to block gene expression. "Ribozyme" refers to a catalytic RNA and includes sequence-specific endoribonucleases.

As used herein, "suitable regulatory sequences" refer to nucleotide sequences in native or chimeric genes that are located upstream (5'), within, and/or downstream (3') to the nucleic acid fragments of the invention, which control the expression of the nucleic acid fragments of the invention. The term "expression", as used herein, refers to the transcription and stable accumulation of the sense (mRNA) or the antisense RNA derived from the nucleic acid fragment(s) of the invention that, in conjunction with the protein apparatus of the cell, results in altered levels of the fatty acid desaturase(s). Expression or overexpression of the gene involves transcription of the gene and translation of the mRNA into precursor or mature fatty acid desaturase proteins. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of preventing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Cosuppression" refers to the expression of a foreign gene which has substantial homology to an endogenous gene resulting in the suppression of expression of both the foreign and the endogenous gene. "Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Promoter" refers to a DNA sequence in a gene, usually upstream (5') to its coding sequence, which controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. In artificial DNA constructs promoters can also be used to transcribe antisense RNA. Promoters may also contain DNA sequences that are involved in the binding of

protein factors which control the effectiveness of transcription initiation in response to physiological or developmental conditions. It may also contain enhancer elements. An "enhancer" is a DNA sequence which can stimulate promoter activity. It may be an innate element of the promoter or a heterologous element inserted to enhance the level and/or tissue-specificity of a promoter. "Constitutive promoters" refers to those that direct gene expression in all tissues and at all times. "Tissue-specific" or "development-specific" promoters as referred to herein are those that direct gene expression almost exclusively in specific tissues, such as leaves or seeds, or at specific development stages in a tissue, such as in early or late embryogenesis, respectively.

The "3' non-coding sequences" refers to the DNA sequence portion of a gene that contains a polyadenylation signal and any other regulatory signal capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor.

The term "Transit Peptide" refers to the N-terminal extension of a protein that serves as a signal for uptake and transport of that protein into an organelle such as a plastid or mitochondrion.

"Transformation" herein refers to the transfer of a foreign gene into the genome of a host organism and its genetically stable inheritance. "Restriction fragment length polymorphism" refers to different sized restriction fragment lengths due to altered nucleotide sequences in or around variant forms of genes. "Fertile" refers to plants that are able to propagate sexually.

"Oil-producing species" herein refers to plant species which produce and store triacylglycerol in specific organs, primarily in seeds. Such species include soybean (Glycine max), rapeseed and canola (including Brassica napus, B. campestris), sunflower (Helianthus annus), cotton (Gossypium hirsutum), corn (Zea mays), cocoa (Theobroma cacao), safflower (Carthamus tinctorius), oil palm (Elaeis guineensis), coconut palm (Cocos nucifera), flax (Linum usitatissimum), castor (Ricinus communis) and peanut (Arachis hypogaea). The group also includes non-agronomic species which are useful in developing appropriate expression vectors such as tobacco, rapid cycling Brassica species, and Arabidopsis thaliana, and wild species which may be a source of unique fatty acids.

"Sequence-dependent protocols" refer to techniques that rely on a nucleotide sequence for their utility. Examples of sequence-dependent protocols include, but are not limited to, the methods of nucleic acid and oligomer hybridization and methods of DNA and RNA amplification such as are exemplified in various uses of the polymerase chain reaction. "PCR product" refers to the DNA product obtained through polymerase chain reaction.

Various solutions used in the experimental manipulations are referred to by their common names such as "SSC", "SSPE", "Denhardt's solution", etc. The composition of these solutions may be found by reference to Appendix B of Sambrook, et al. (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press).

T-DNA Mutagenesis and Identification of an
Arabidopsis Mutant Defective in Delta-15 Desaturation

In T-DNA mutagenesis (Feldmann, et al., Science (1989) 243:1351-1354), the integration of T-DNA in the genome can interrupt normal expression of the gene at or near the site of the integration. If the resultant mutant phenotype can be detected and shown genetically to be tightly linked to the T-DNA insertion, then the "tagged" locus and its wild type counterpart can be readily isolated by molecular cloning by one skilled in the art.

Arabidopsis thaliana seeds were transformed by *Agrobacterium tumefaciens* C58C1rif strain harboring the avirulent Ti-plasmid pGV3850::pAK1003 that has the T-DNA region between the left and right T-DNA borders replaced by the origin of replication region and ampicillin resistance gene of plasmid pBR322, a bacterial kanamycin resistance gene, and a plant kanamycin resistance gene (Feldmann, et al., Mol. Gen. Genetics (1987) 208:1-9). Plants from the treated seeds were self-fertilized and the resultant progeny seeds, germinated in the presence of kanamycin, were self-fertilized to give rise to a population, designated T3, that was segregating for T-DNA insertions. T3 seeds from approximately 6000 T2 plants were analyzed for fatty acid composition. One line, designated 3707, showed a reduced level of linolenic acid (18:3). One more round of self-fertilization of mutant line 3707 produced T4 progeny seeds. The ratio of 18:2/18:3 in seeds of the homogyzous mutant in T4 population was ca. 14; this ratio is ca 1.8 and ca. 23, respectively, in wild-type Arabidopsis and Arabidopsis fad 3 mutant [Lemieux et al. (1990) Theor. App. Gen. 80:234-240] obtained via chemical mutagenesis. These seeds were planted and 263 individual plants were analyzed for the presence of

nopaline in leaf extracts. T5 seeds from these plants were furth r analyzed for fatty acid composition and the ability to germinate in the presence of kanamycin. The mutant fatty acid phenotype was found to segregate in a 1:2:1 ratio, as was germinability on kanamycin. Nopaline was found in all plants with an altered fatty acid phenotype, but not in wild type segregants. These results provided evidence that the locus controlling delta-15 desaturation was interrupted by T-DNA in mutant line 3707.

Isolation of Arabidopsis Genomic DNA
Containing the Gene Controlling Delta-15 Desaturation

In order to isolate the gene controlling delta-15 desaturation from wild-type Arabidopsis, a T-DNA-plant DNA "junction" fragment containing a T-DNA border integrated into the host plant DNA was isolated from Arabidopsis mutant 3707. For this, genomic DNA from the mutant plant was isolated and completely digested by either Bam HI or Sal I restriction enzymes. In each case, one of the resultant fragments was expected to contain the origin of replication and ampicillin-resistance gene of pBR322 as well as the left T-DNA-plant DNA junction fragment. Such fragments were rescued as plasmids by ligating the digested genomic DNA fragments at a dilute concentration to facilitate self-ligation and then using the ligated fragments to transform E. coli cells. Ampicillin-resistant E. coli transformants were isolated and screened by colony hybridization to fragments containing either the left or the right T-DNA border. Of the 192 colonies obtained from the plasmid rescue of Sal I digested genomic DNA, 31 hybridized with the left T-DNA border fragment, 4 hybridized to the right T-DNA border fragment, and none hybridized to both. Of the 85 colonies obtained from the plasmid rescue of Bam HI digested genomic DNA, 63

hybridized to the left border and none to the right border. Restriction analysis of seven rescued plasmids that were obtained from the Bam HI digestion and that hybridized to the left T-DNA border showed that they were indistinguishable and contained 1.4 kb of putative, flanking plant DNA. Restriction analysis of another rescued plasmid, pS1, that was obtained from the Sal I digestion and hybridized only to the left T-DNA border, showed that it contained 2.9 kb of putative, flanking plant DNA. This flanking DNA had a Bam HI site and a Hind III site 1.4 kb and 2.2 kb, respectively, away from the left T-DNA border, suggesting that the 1.4 kb putative plant DNA in Bam HI rescued plasmids was contained within the 2.9 kb putative plant DNA in the Sal I rescued plasmids. Southern blot analysis of wild type and mutant 3707 Arabidopsis genomic DNA using the radiolabeled 1.4 kb DNA fragment as the hybridization probe confirmed that this fragment contained plant DNA and that the T-DNA integration site was in a 2.8 kb Bam HI, a 5.2 kb Hind III, a 3.5 kb Sal I, a 5.5 kb Eco RI, and an approximately 9 kb Cla I fragment of wild type Arabidopsis DNA. Nucleotide sequencing of plasmid pS1 with a primer made to a left T-DNA border sequence revealed that pS1 was colinear with the sequence of the left T-DNA border (Yadav et al., Proc. Natl. Acad. Sci. USA (1982) 79:6322-6326) up to nucleotide position 65, which is in the T-DNA border repeats. Approximately 800 bp of additional sequence in pS1 beyond the T-DNA-plant DNA junction, that is, in the plant DNA adjoining the left T-DNA border, showed no significant homology to the T-DNA of pGV3850::pAK1003 and no significant open reading frame.

The nucleic acid fragment from wild-type Arabidopsis corresponding to the plant DNA flanking T-DNA in the line 3707 was isolated by screening a

lambda phage _Arabidopsis thaliana_ genomic library with the 1.4 kb plant DNA isolated from the rescued plasmids as a hybridization probe. Seven positively-hybridizing genomic clones were isolated that fell in one of five classes based on partial restriction mapping. While their average insert size was approximately 15 kb, taken together they spanned a total of approximately 40 kb of genomic DNA. A combination of restriction and Southern analyses revealed that the five clones overlapped the site of integration of the left border of the T-DNA and that there was no detectable rearrangement of plant DNA in the rescued plasmids as compared to that in the wild type genomic plant DNA. One of these lambda phage clones, designated 1111, was representative of the recovered clones and contained an approximately 20 kb genomic DNA insert which was more or less symmetrically arranged around the site of insertion of the left border of the T-DNA. This clone was deposited on November 27, 1991 with the American Type Culture Collection of Rockville, Maryland under the provisions of the Budapest Treaty and bears accession number ATCC 75167.

Isolation of _Arabidopsis_ Delta-15
Desaturase cDNA

A 5.2 kb Hind III fragment containing wild-type genomic DNA, which hybridized to the 1.4 kb flanking plant DNA recovered from line 3707 and which was interrupted near its middle by the T-DNA insertion in line 3707, was isolated from lambda phage clone 41A1 and cloned into the Hind III site of the pBluescript SK vector (Stratagene) by standard cloning procedures described in Sambrook et·al., Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press). The resultant plasmid was designated pF1. The isolated 5.2 kb Hind III fragment was also used as a radiolabeled hybridization probe to screen a

cDNA library made to poly A^+ mRNA from 3-day-old etiolated Arabidopsis thaliana (ecotype Columbia) seedling hypocotyls in a lambda ZAP II vector (Stratagene). Of the several positively-hybridizing plaques, four strongly-hybridizing ones were subjected to plaque purification. Sequences of the pBluescript (Stratagene) vector, including the cDNA inserts, from each of the purified phage stocks were excised in the presence of a helper phage. The resultant phagemids were used to infect E. coli cells which yielded double-stranded plasmids, pCF1, pCF2, pCF3, and pCF4. All four were shown to contain at least one approximately 1.3 to 1.4 kb Not I insert fragment (Not I/Eco RI adaptors were used in the preparation of the cDNA library) which hybridized to the same region of wild-type plant genomic DNA present in the isolated phage clones. This region, which was near the site of integration of the left T-DNA border in line 3707, was on the side of the T-DNA insertion opposite to that of the plant DNA flanking the left T-DNA border isolated previously via plasmid rescue. Partial sequence determination of the different cDNAs revealed common identity. Since multiple versions of only one type of cDNA were obtained from a cDNA library made from etiolated tissue which is expected to express delta-15 desaturation, and since these cDNAs hybridized to the genomic DNA that corresponds to the site of T-DNA integration in line 3707 which had a high linoleic acid/low linolenic acid phenotype, Applicants were lead to conclude that the T-DNA in line 3707 interrupted the normal expression of the gene encoding delta-15 desaturase. The complete nucleotide sequence of one cDNA, designated pCF3, was determined and is shown as SEQ ID NO:1. It reveals an open reading frame that encodes a 386 amino acid polypeptide. One of the sequencing primers made to the pCF3 insert was also used

to obtain 255 bp of sequence from pF1 that is shown as SEQ ID NO:3. Nucleotides 68 to 255 of the genomic DNA in pF1 (SEQ ID NO:3) are identical to nucl otides 1 to 188 of the cDNA (SEQ ID NO:1), which shows that they are colinear and that the cDNA is encoded for by the gene in the isolated genomic DNA. Nucleotides 113 to 115 in SEQ ID NO:3 are the initiation codon of the largest open reading frame corresponding to nucleotides 46-48 in SEQ ID NO:1. This is evident from the presence of in-frame termination codons at nucleotides 47 to 49 and nucleotides 56 to 58 and the absence of observable intron splice junctions in SEQ ID NO:3. The identification of the 386 amino acid polypeptide as a desaturase was confirmed by comparing its amino acid sequence with all the protein sequences found in Release 19.0 of the SWISSPROTEIN database using the FASTA algorithm of Pearson and Lipman (Proc. Natl. Acad. Sci. USA (1988) 85:2444-2448) and the BLAST program (Altschul et al., J. Mol. Biol. (1990) 215:403-410). The most homologous protein found in both searches was the desA fatty acid desaturase from the cyanobacterium Synechocystis PCC6803 (Wada, et al., Nature (1990) 347:200-203; Genbank ID:CSDESA; GenBank Accession No:X53508). The 386 amino acid peptide in SEQ ID NO:1 was also compared to the 351 amino acid sequence of desA by the method of Needleman et al. (J. Mol. Biol. (1970) 48:443-453). Over their entire length, these proteins were 26% identical, the comparison imposing four major gaps in the desA protein sequence. While this overall homology is poor, homology in shorter stretches was better. For instance, in a stretch of 78 amino acids the Arabidopsis delta-15 desaturase (amino acids 78 to 155 in SEQ ID NO:1) and the desA protein (amino acids 67 to 144) showed 40% identity and 66% similarity.

Homology in yet shorter stretches was even greater as shown in Table 2.

TABLE 2

Peptide Length	AA positions in SEQ ID NO:1	AA positions in desA	Percent Identity
12	97-108	86-97	83
7	115-121	104-110	71
9	133-141	22-130	56
11	299-309	282-292	64

These high percent identities in short stretches of amino acids between the cyanobacterial desaturase polypeptide and SEQ ID NO:2 suggests significant relatedness between the two.

To analyse the developmental expression of the gene encoding mRNA coresponding to SEQ ID NO:1, the cDNA insert in plasmid pCF3 was used as a radiolabeled hybridization probe on mRNA samples from leaf, root, germinating seedling, and developing siliques from both wild type amd mutant 3707 Arabidopsis plants, essentially as described in Maniatis et al., Molecular Cloning, A Laboratory Manual (1982) Cold Spring Harbor Laboratory Press. The results indicated that while the mRNA corresponding to SEQ ID NO:1 is detected in all tissues from the mutant plant, its levels are lower than in wild-type tissues. This is consistent with the observation that the fatty acid mutation in line 3707 is leaky relative to the known Arabidopsis fad 3 mutant obtained via chemical mutagenesis. These results confirmed that the T-DNA in line 3707 had interrupted the normal expression of a fatty acid desaturase gene. Based on the fatty acid phenotype of homozygous mutant line 3707, Applicants concluded that the cDNA insert in pCF3 encoded the delta-15 desaturase. Further, Applicants concluded that it was the microsomal delta-15 desaturase, and not the chloroplastic delta-15

desaturase, since: a) the mutant phenotype was expressed strongly in the seed but express d poorly, if at all, in the leaf of line 3707, and b) the delta-15 desaturase polypeptide, by comparison to the desA polypeptide, did not have an N-terminal extension of a transit peptide expected for a nuclear-encoded chloroplast desaturase.

The identity of SEQ ID NO:2 as the *Arabidopsis* microsomal delta-15 desaturase was confirmed by its biological overexpression in plant tissues. For this, the 1.4 kB Not I fragment of plasmid pCF3 containing the delta-15 desaturase cDNA was placed in the sense orientation behind either the CaMV 35S promotor, to provide constituitive expression, or behind the promotor for the gene encoding soybean a' subunit of the β-conglycinin (7S) seed storage protein, to provide embryo-specific expression. The chimeric genes 35S promoter/sense SEQ ID NO:1/3' nopaline synthase and β-conglycinin/sense SEQ ID NO:1/3' phaseolin were then transformed into plant cells by *Agrobacterium tumefaciens*'s binary Ti plasmid vector system [Hoekema et al. (1983) Nature 303:179-180; Bevan (1984) Nucl. Acids Res. 12:8711-8720].

To confirm the identity of SEQ ID NO:1 and to test the biological effect of its overexpression in a heterologous plant species, the chimeric genes 35S promoter/sense SEQ ID NO:1/3' nopaline synthase was transformed into a binary vector, which was then transferred into *Agrobacterium tumefaciens* strain R1000, carrying the Ri plasmid pRiA4b from *Agrobacterium rhizogenes* [Moore et al. (1979) Plasmid 2:617-626]. Carrot (*Daucus carota* L.) cells were transformed by co-cultivation of carrot root disks with strain R1000 carrying the chimeric gene by the method of Petit et al. (1986) [Mol. Gen. Genet. 202:388-393]. Fatty acid

analyses of transgenic carrot "hairy" roots show that overexpression of Arabidopsis microsomal d lta-15 desaturase can result in over 10-fold increase in 18:3 at the expense of 18:2.

To complement the delta-15 desaturation mutation in the T-DNA mutant line 3707 and to test the biological effect of overexpression of SEQ ID NO:1 in seed, the embryo-specific promoter/SEQ ID NO:1/3' phaseolin chimeric gene was transformed into a binary vector, which was then transformed into the avirulent Agrobacterium strain LBA4404/pAL4404 [Hoekema et al. (1983) Nature 303:179-180]. Roots of line 3707 were transformed by the engineered Agrobacterium, transformed plants were selected and grown to give rise to seeds. Fatty acid analysis of the seeds from two plants showed that the one out of six seeds in each plant showed the mutant fatty acid phenotype, while the remaining seeds show more than 10-fold increase in 18:3 to ca. 55%. While the sample size is small, this segregation suggests Mendelian inheritance of the fatty acid phenotype. While most of the increase occurs at the expense of 18:2, some of it also occurs at the expense of 18:1. Thus, overexpression of this gene in oils crops, especially canola, which is a close relative of Arabidopsis, is also expected to result in the high levels of 18:3 that are found in specialty oil of linseed.

Comparisons of the sequence of the 386 amino acid polypeptide by the method of Needleman et al. (J. Mol. Biol. (1970) 48:443-453) with those for the microsomal stearoyl-CoA (delta-9) desaturases from rat, mouse and yeast revealed 21%, 19%, and 17% identities, respectively. While the membrane-associated Arabidopsis delta-15 desaturase protein showed significant but limited homology to the desA protein, it showed no

significant homology to the soluble stearoyl-ACP (delta-9) desaturases from higher plants, including one from Arabidopsis.

Comparison of partial nucleotide sequences of plasmids pF1 and pS1 showed that the left T-DNA border:plant DNA junction is ca. 700 bp from the initiaton codon in SEQ ID NO:1. To determine the position of the other T-DNA:plant DNA junction with respect to the pF1 sequence, the T-DNA:plant DNA junction fragment was isolated. Genomic DNA from mutant line 3707, isolated as described previously, was partially digested by restriction enzyme Mbo I to give an average fragment size of ca. 15 kB. The fragment ends were partially-filled with dGTP and gATP by Klenow and cloned into Xho I half-sites of LambdaGEM®-11 (Promega Corporation) following the manufacturer's protocol. The phage library was titered and used essentially as described in Ausubel et al. [Current Protocols in Molecular Biology (1989) John Wiley & Sons]. The genomic phage library was screened with radiolabeled PCR product, ca. 0.6 kB, derived from 5' end of the gene in pF1. This product spans from 3 bp to the right of where the left-T-DNA border inserted to 15 bp to the left of nucleotide position 1 in SEQ ID NO:1. Southern blot analysis of DNA from one of the purified, positively-hybridizing phages following Eco RI restriction digestion and electrophoresis showed that a 4 kB Eco RI fragment hybridized to the 0.6 kB PCR product. The Eco RI fragment was subcloned and subject to sequence analyses. Comparison of the sequences derived from this fragment, pF1 and pS1 showed that the insertion of T-DNA resulted in a 56 bp deletion at the site of insertion and that the T-DNA interrupted the Arabidopsis gene 711 bp 5' to the initiaton codon in SEQ ID NO:1. Thus, the T-DNA inserts 5' to the open reading

frame, consistent with the leaky expresssion of the gene encoding SEQ ID NO:1 and the leaky fatty acid phenotype in mutant 3707. While the left T-DNA:plant DNA junction is precise, that is without any sequence rearrangement in either the left T-DNA border or the flanking plant DNA, the other T-DNA:plant DNA junction is complex and not fully characterized.

Plasmid pCF3 was deposited on December 3, 1991 with the American Type Culture Collection of Rockville, Maryland under the provisions of the Budapest Treaty and bears accession number ATCC 68875.

Using _Arabidopsis_ Delta-15 Desaturase cDNA as a Hybridization Probe to Isolate cDNAs Encoding _Related Desaturases from Arabidopsis_

The 1.4 kb Not I insert fragment isolated from plasmid pCF3 was purified, radiolabeled, and used to screen approximately 80,000 clones from the cDNA library made to poly A^+ mRNA from 3-day-old etiolated _Arabidopsis thaliana_ as described above, except that lower stringency hybridizations (1 M NaCl, 50 mM Tris-HCl, pH 7.5, 1% SDS, 5% dextran sulfate, 0.1 mg/mL denatured salmon sperm DNA and 50°C) and washes (sequentially with 2X SSPE, 0.1% SDS at room temperature for 5 min and then again with fresh solution for 10 min, and finally with 0.5X SSPE, 0.1% SDS at 50°C for 5 min.) were used. Approximately 17 strongly-hybridizing and 17 weakly-hybridizing plaques were identified in the primary screen. Four of the weakly-hybridizing plaques were picked and subjected to one or two further rounds of screening with the radiolabeled probe as above until they were pure. To ensure that these were not delta-15 desaturase clones, they were further analyzed to determine whether they hybridized to an 18 bp oligomer specific to the 3' non-coding region of delta-15 desaturase cDNA (pCF3). After autoradiography of the

filters, one of the clones was found not to hybridize to this probe. This clone was picked, and a plasmid clone containing the cDNA insert was obtained as described above. Restriction analysis of this plasmid, designated pCM2, showed that it had an approximately 1.3 kb cDNA insert which lacked a 0.7 kb Nco I - Bgl II fragment characteristic of the _Arabidopsis_ delta-15 desaturase cDNA of pCF3. (This fragment corresponds to the DNA located between the Nco I site at nucleotides 474 to 479 and the Bgl II site at nucleotides 1164 to 1169 in SEQ ID NO:1). Partial nucleotide sequences of single strands from the 5' region and 3' region of pCM2 revealed that the cDNA insert was incomplete and that it encoded a polypeptide that is similar to, but distinct from, that encoded by the cDNA in pCF3. In order to isolate a full-length version of the cDNA in plasmid pCM2, the 1.3 kB Not I fragment from plasmid pCM2 containing the cDNA insert was isolated and used as a radiolabeled hybridization probe to rescreen the same _Arabidopsis_ cDNA library as above. Three strongly hybridizing plaques were purified and the plasmids excised as described previously. The three resultant plasmids were digested by Not I restriction enzyme and shown to contain cDNA inserts ranging in size between 1 kB and 1.5 kB. Complete nucleotide sequence determination of the cDNA insert in one of these plasmids, designated pACF2-2, is shown in SEQ ID NO:4. SEQ ID NO:4 shows the 5' to 3' nucleotide sequence of base pairs of the _Arabidopsis thaliana_ cDNA which encodes a fatty acid desaturase. Nucleotides 10-12 and nucleotides 1358 to 1350 are, respectively, the putative initiation codon and the termination codon of the open reading frame (nucleotides 10 to 1350). The open reading frame was confirmed by comparison of its deduced amino acid sequences with that of the related delta-15

fatty acid desaturase from soybean in this application. Nucleotides 1 to 9 and 1351 to 1525 are, respectively, the 5' and 3' untranslated nucleotides. The 446 amino acid protein sequence in SEQ ID NO:5 is that deduced from the open reading frame in SEQ ID NO:4 and has an estimated molecular weight of 51 kD. Alignment of SEQ ID NOS:2 and 5 shows an overall homology of approximately 80% and that the former has an approximately 55 amino acid long N-terminal extension, which is deduced to be a transit peptide found in nuclear-encoded plastid proteins.

To analyse the developmental expression of the gene corresponding to SEQ ID NO:4, this sequence was used as a radiolabeled hybridization probe on mRNA samples from leaf, root, germinating seedling, and developing siliques from both wild type and mutant line 3707 Arabidopsis plants, essentially as described in Maniatis et al. [Molecular Cloning, A Laboratory Manual (1982) Cold Spring Harbor Laboratory Press]. The results indicated that, in contrast to the constitutive expression of the gene encoding SEQ ID NO:1, the mRNA corresponding to SEQ ID NO:4 is abundant in green tissues, rare in roots and leaves, and is about three-fold more abundant in leaf than that of SEQ ID NO:1. The cDNA in plasmid pCM2 was also shown to hybridize polymorphically to genomic DNA from Arabidopsis thaliana (ecotype Wassileskija and marker line W100 ecotype Landesberg background) digested with Eco RI. It was used as a RFLP marker to map the genetic locus for the gene encoding this fatty acid desaturase in Arabidopsis. A single genetic locus was positioned corresponding to this desaturase cDNA. Its location was thus determined to be on chromosome 3 between the lambda AT228 and cosmid c3838 RFLP markers, "north" of the glabrous locus (Chang et al., Proc. Natl. Acad. Sci. USA (1988)

85:6856-6860; Nam et al., Plant Cell (1989) 1:699-705). This approximates the region to which Arabidopsis fatty acid desaturase fad 2, fad D, and fad B mutations map [Somerville et al., (1992) in press]. Unsuccessful efforts to clone the microsomal delta-12 fatty acid desaturase using cDNA inserts of pCF3 and pACF2-2 alongwith the above data led Applicants to conclude that the cDNA in pACF2-2 encodes a plastid delta-15 fatty acid desaturase that corresponds to the fad D locus. This conclusion will be confirmed by biological expression of the cDNA in pACF2-2.

Plasmid pCM2 was deposited on November 27, 1991 with the American Type Culture Collection of Rockville, Maryland under the provisions of the Budapest Treaty and bears accession number ATCC 68852.

The 1.4 kb, 1.3 kB, and 1.5 kB Not I cDNA insert fragments isolated from plasmids pCF3, pCM2 and pACF2-2 were purified, radiolabeled, and used several times to screen at low stringency as described above two different cDNA libraries: one was made to poly A^+ mRNA from 3-day-old etiolated Arabidopsis thaliana ("etiolated" library) as described above and one made to polyA^+ mRNA from the above-ground parts of Arabidopsis thaliana plants, which varied in size from those that had just opened their primary leaves to plants which had bolted and were flowering [Elledge et al. (1991) Proc. Natl. Acad Sci. USA 88:1731-1735]. The cDNA inserts in the library were made into an Xho I site flanked by Eco RI sites in lambda Yes vector [Elledge et al. (1991) Proc. Natl. Acad Sci. USA 88:1731-1735] ("leaf" library). Several plaques from both libraries that hybridized weakly and in duplicate lifts to both SEQ ID NOS:1 and 4 were subjected to plaque purification. Phagemids were excised from the pure phages from "etiolated" library as described above. Plasmids were

excised from the purified phages of the "leaf" library by site-specific recombination using the cre-lox recombination system in E. coli strain BNN132 [Elledge et al. (1991) Proc. Natl. Acad Sci. USA 88:1731-1735]. In all cases, nucleotide sequencing of the cloned DNA revealed clones either identical to SEQ ID NOS:1 or 4 or unrecognizable sequences.

In another set of experiments ca. 400,000 phages in the "leaf" library was screened with SEQ ID NOS:1 and 4 at low stringency (26 C, 1 M Na+, 50% formamide) and high stringency (42 C, 1 M Na+, 50% formamide). Of the several positive signals on the primary plaque lifts, 11 showed high stringency hybridization to SEQ ID NO:1, 35 showed high stringency hybridization to SEQ ID NO:4, and 39 hybridized to both at low stringency only. Twenty seven plaques of the low stringency signals came through a secondary low-stringency screen, 17 of which were used to make DNA from excised plasmids. Of the 7 plasmid DNA were sequenced, 8 were unrecognizable sequences, 5 were identical to SEQ ID NO:1, 2 were identical to SEQ ID NO:2, and 2 were identical to one another and related but distinct to SEQ ID NOS:1 and 4. The novel desaturase sequence, designated pFad-x2, was also isolated from the "leaf" library independently by using as a hybridization probe a 0.6 kB PCR product derived by polymerase chain reaction on poly A+ RNA made from both canola seed as well as Arabidopsis leaves, as described elsewhere in this application, using degenerate oligomers made to conserved sequences between plant delta-15 desaturases and the cyanobacterial des A desaturase. The PCR-derived plasmid, designated pYacp7, was sequenced partially from both ends. Comparison of the sequences of pFad-x2 and pYacp7 revealed that the two independently cloned cDNAs contained an identical sequence that was related to the other delta-15

desaturases and that both were incomplete cDNAs. A partial composite sequence derived from both plasmids, pFadx-2 and pYacp7, is shown in SEQ ID NO:16 as a 5' to 3' nucleotide sequence of 472 bp. Nucleotides 2-4 and nucleotides 468 to 470 are, respectively, the first and the last codons in the open reading frame. This open reading frame is shown in SEQ ID NO:17. Comparison of SEQ ID NO:17 to the other delta-15 desaturase polypeptides disclosed in this application by the method of Needleman et al. [J. Mol. Biol. (1970) 48:443-453)] using gap weight and gap length weight values of 3.0 and 0.1, respectively. The overall identities are between 65% and 68% between SEQ ID NO:17 and the microsomal delta-15 desaturases from Arabidopsis, canola and soybean and the overall identities are between 77% and 87% between SEQ ID NO:17 and the plastid delta-15 desaturases from Arabidopsis, canola and soybean. In addition SEQ ID NO:17 has an N-terminal peptide extension compared to the microsomal delta-15 desaturases that shows homology of the transit peptide sequence in Arabidopsis plastid delta-15 desaturase. On the basis of these comparisons it is deduced that SEQ ID NO:16 encodes a plastid delta-15 desaturase. There is genetic data in Arabidopsis suggesting the presence of two loci for plastid delta-15 desaturase. The full-length version of SEQ ID NO:16 can be readily isolated by one skilled in the art. The biological effect of introducing SEQ ID NO:16 or its full-length version into plants will be used to confirm its identity.

Plasmid pYacp7 was deposited on 20 November 1992 with the American Type Culture Collection of Rockville, Maryland under the provisions of the Budapest Treaty and bears accession number ATCC 69129.

Using *Arabidopsis* Delta-15 Desaturase cDNAs as Hybridization Probes to Isolate Delta-15 Desaturase cDNAs from Other Plant Species

For the purpose of cloning the *Brassica napus* seed cDNAs encoding delta-15 fatty acid desaturases, the cDNA inserts from pCF3 and pCM2 were isolated by polymerase chain reaction from the respective plasmids, radiolabeled, and used as hybridization probes to screen a lambda phage cDNA library made with poly A^+ mRNA from developing *Brassica napus* seeds 20-21 days after pollination. This cDNA library was screened several times at low stringency, using the *Arabidopsis* cDNA probes mentioned above. One of the *Brasssica napus* cDNAs obtained in the initial screens was used as probe in a subsequent high stringency screen.

Arabidopsis pCM2 insert was radiolabeled and used as probe to screen approximately 300,000 plaques under low stringency hybridization conditions. The filter hybridizations were performed in 50 mM Tris pH 7.6, 6X SSC, 5X Denhardt's, 0.5% SDS, 100 ug denatured calf thymus DNA at 50°C overnight, and the posthybridization washes were carried out in 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. Five strongly-hybridizing phages were obtained. These were plaque purified and used to excise the phagemids as described in the manual of the pBluescriptII Phagemid Kit from Stratagene (Stratagene 1991 catalogue, item 212205). One of these, designated pBNSF3-2, contained a 1.3 kb insert. pBNSF3-f2 was sequenced completely on both strands and the nucleotide sequence is shown in SEQ ID NO:6. Plasmid pBNSF3-2 was deposited on 27 November 1991 with the American Type Culture Collection of

Rockville Maryland, USA under the provisions of the Budapest Treaty and bears the accession number 68854.

An additional low stringency screen using pCM2 probe provided eight strongly hybridizing phages. One of these, designated pBNSFd 8, contained a 0.4kb insert. pBNSFd-8 was sequenced completely on one strand, this nucleotide sequence showed significant divergence from the sequence SEQ ID NO:6 in the homologous region, which suggested that it corresponded to a novel Brassica napus seed desaturase different from that shown in SEQ ID NO:6. pBNSFd-8 insert was radiolabelled and used as hybridization probe in a high stringency screen of the Brassica napus seed cDNA library. The hybridization conditions were identical to those of the low stringency screen described above except for the temperature of the final two 30 min posthybridization washes in 0.2x SSC, 0.5% SDS was increased to 60°C. This screen resulted in three strongly hybridizing phages that were purified and excised. One of the excised plasmids pBNSFd-3 contained a 1.4kb insert that was sequenced completely on both strands. SEQ ID NO:8 shows the complete nucleotide sequence of pBNSFd-2.

Using Arabidopsis Delta-15 Desaturase cDNA as a Hybridization Probe to Isolate a Glycerolipid Desaturase cDNA from Soybean

A cDNA library was made to poly A^+ mRNA isolated from developing soybean seeds, and screened essentially as described above, except that filters were prehybridized in 25 mL of hybridization buffer consisting of 50mM Tris-HCl, pH 7.5, 1 M NaCl, 1% SDS, 5% dextran sulfate and 0.1 mg/mL denatured salmon sperm DNA (Sigma Chemical Co.) at 50°C for 2 h. Radiolabeled probe prepared from pCF3 as described above was added, and allowed to hybridize for 18 h at 50°C. The probes were washed twice at room temperature with 2X SSPE, 1%

SDS for five min followed by washing for 5 min at 50°C in 0.2X SSPE, 1% SDS. Autoradiography of the filt rs indicated that there was one strongly hybridizing plaque, and approximately five weakly hybridizing plaques. The more strongly hybridizing plaque was subjected to a second round of screening as before, except that the final wash was for 5 min at 60°C in 0.2X SSPE, 1% SDS. Numerous, strongly hybridizing plaques were observed, and one, well-isolated from other phage, was picked for further analysis.

Sequences of the pBluescript vector from the purified phage, including the cDNA insert, were excised in the presence of a helper phage and the resultant phagemid was used to infect E. coli XL-1 Blue cells. DNA from the plasmid, designated pXF1, was made by the alkaline lysis miniprep procedure described in Sambrook et al. (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989) Cold Spring Harbor Laboratory Press). The alkali-denatured double-stranded DNA from pXF1 was completely sequenced on both strands. The insert of pXF1 contained a stretch of 1783 nucleotides which contained an unknown open-reading frame and also contained a poly-A stretch of 16 nucleotides 3' to the open reading frame, from nucleotides 1767 to 1783, followed by an Eco RI restriction site. The 2184 bases that followed this Eco RI site contained a 1145 bp open reading frame which encoded a polypeptide of about 68% identity to, and colinear with, the Arabidopsis delta-15 desaturase polypeptide listed in SEQ ID No:2. The putative start methionine of the 1145 bp open-reading frame corresponded to the start methionine of the Arabidopsis microsomal delta-15 peptide and there were no amino acids corresponding to a plastid transit peptide 5' to this methionine. When the insert in pXF1 was digested with Eco RI four fragments were observed,

fragments of approximately 370 bp and 1400 bp fragments, derived from the first 1783 bp of the insert in pXF1, and fragments of approximately 600 bp and 1600 bp derived from the the other 2184 nucleotides of the insert in pXF1. Only the 600 bp and 1600 bp fragments hybridized with probe derived from pCF3 on Southern blots. It was deduced that pXF1 contained two different cDNA inserts separated by an Eco RI site and the second of these inserts was a 2184 bp cDNA encoding a soybean microsomal delta-15 desaturase. The complete nucleotide sequence of the 2184 bp soybean microsomal delta-15 cDNA contained in plasmid pXF1 is listed in SEQ ID No:10. Plasmid pXF1 was deposited on December 3, 1991 with the American Type Culture Collection of Rockville, Maryland under the provisions of the Budapest Treaty and bears accession number ATCC 68874.

Using Soybean Microsomal Delta-15 Desaturase cDNA as a Hybridization Probe to Isolate cDNAs Encoding Related Desaturases from Soybean

A 1.0 kb fragment of DNA corresponding to part of the coding region of the soybean microsomal delta-15 desaturase cDNA contained in plasmid pXF1, was excised with the restriction enzyme Hha I and gel purified. The fragment was labeled with ^{32}P as described above and used to probe a soybean cDNA library as described above. Autoradiography of the filters indicated that there were eight hybridizing plaques and these were subjected to a second round of screening. Sequences of the pBluescript vector from all eight of the purified phages, including the cDNA inserts, were excised in the presence of a helper phage and the resultant phagemids were used to infect E. coli XL-1 Blue cells. DNA from the plasmids was made by the alkaline lysis miniprep procedure described in Sambrook et al. (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989) Cold Spring Harbor

Laboratory Press). Restriction analysis showed they contained inserts ranging from 1.0 kb to 3.0 kb in size. One of these inserts, designated pSFD-118bwp, contained an insert of about 1700 bp. The alkali-denatured double-stranded DNA from pSFD-118bwp was completely sequenced on both strands, shown in SEQ ID NO:12. The insert of pSFD-118bwp contained a stretch of 1675 nucleotides which contained an open-reading frame encoding a polypeptide, shown in SEQ ID NO:13, of about 80% identity with, and colinear with, the Arabidopsis plastid delta-15 desaturase polypeptide listed in SEQ ID No:5. The open-reading frame also encoded amino acids corresponding to a plastid transit peptide at the 5' end of the open-reading frame. The transit peptide was colinear with, and shared some homology to, the transit peptide described for the Arabidopsis plastid delta-15 glycerolipid desaturase. The complete nucleotide sequence of the 1675 bp soybean plastid delta-15 glycerolipid desaturase cDNA is listed in SEQ ID No:12.

Comparison of the different delta-15 desaturase sequences disclosed in the application by the method of Needleman et al. (J. Mol. Biol. (1970) 48:443-453) using gap weight and gap length weight values of 3.0 and 0.1, respectively, reveals the relatedness between them as shown in Table 3.

TABLE 3

Percent Identities Between Different Delta-15 Fatty Acid Desaturases at the Amino Acid Level

	aD	c3	cD	s3	sD
a3	66	93	66	68	67
aD	-	67	90	67	69
c3	-	-	68	68	68
cD	-	-	-	68	74

a3, ad, c3, cD, s3 and sD refer, respectively, to SEQ ID NO:2 (*Arabidopsis* microsomal delta-15 desaturase), SEQ ID NO:5 (*Arabidopsis* plastid delta-15 desaturase), SEQ ID NO:7 (canola microsomal delta-15 desaturase), SEQ ID NO:9 (canola plastid delta-15 desaturase), SEQ ID NO:11 (soybean microsomal delta-15 desaturase), and SEQ ID NO:13 (soybean plastid delta-15 desaturase). Based on these comparisons, the delta-15 desaturases, of both microsomal and plastid types, have overall identities of 65% or more at the amino acid levels, even when from different plant species.

Isolation of Nucleotide Sequences Encoding Homologous and Heterologous Glycerolipid Desaturases

Fragments of the instant invention may be used to isolate cDNAs and genes of homologous and heterologous glycerolipid desaturases from the same species as the fragment of the invention or from different species. Isolation of homologous genes using sequence-dependent protocols is well-known in the art. Southern blot analysis revealed that the *Arabidopsis* microsomal delta-15 desaturase cDNA (SEQ ID NO:1) hybridized to genomic DNA fragments of corn and soybean. In addition, Applicants have demonstrated that it can be used to isolate cDNAs encoding seed microsomal delta-15 desaturases from *Brassica napus* (SEQ ID NO:6) and soybean (SEQ ID NO:10). Thus, one can isolate cDNAs and

genes for homologous glycerolipid desaturases from the same or different higher plant species, especially from the oil-producing species.

More importantly, one can use the fragments of the invention to isolate cDNAs and genes for heterologous glycerolipid desaturases, including those found in plastids. Thus, Arabidopsis microsomal delta-15 desaturase cDNA (SEQ ID NO:1) was successfully used as a hybridization probe to isolate cDNAs encoding the related plastid delta-15 desaturases from Arabidopsis (SEQ ID NO:4) and Brassica napus (SEQ ID NO: 8), and the soybean microsomal delta-15 soybean (SEQ ID NO:10) was successfully used to isolate soybean cDNA encoding plastid delta-15 desaturase (SEQ ID NO:12).

In a particular embodiment of the present invention, regions of the nucleic acid fragments of the invention that are conserved between different desaturases may be used by one skilled in the art to design a mixture of degenerate oligomers for use in sequence-dependent protocols aimed at isolating nucleic acid fragments encoding other homologous or heterologous glycerolipid desaturase cDNA's or genes. For example, by comparing all desaturase polypeptides one can identify stretches of amino acids that are conserved between them, and then use the conserved amino acid sequence to design oligomers, both short degenerate or long ones, or "guessmers" as known by one skilled in the art (see Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press). Such oligomers and "quessmers" may be used as hybridization probes as known to one skilled in the art.

For example, comparison of cyanobacterial desA and plant delta-15 desaturases revealed a particularly well conserved stretch of amino acids (amino acids 97-108 in

SEQ ID NO:1). SEQ ID NOS:20 and 21 represent two sets of 36-mers each 16-fold degenerate made to this r gion. End-labeled oligomers represented in SEQ ID NOS:20 and 21 were mixed and used as hybridzation probes to screen Arabidopsis cDNA libraries. Most of the positively-hybridizing plaques also hybridized to cDNAs encoding Arabidopsis microsomal and plastid delta-15 desaturases (SEQ ID NOS:1 and 4). However, the use of SEQ ID NOS:20 and 21 did not give consistent and reproducible results. A 135 base-long oligomer (SEQ ID NO:32) was also made as an antisense strand to a longer stretch of the same conserved region, amino acids 97 to 141 in SEQ ID NO:1 (FVLGHDCGHGSFSDIPLLNSVVGHILHSFILVPYHGWRISHRTHH). At positions of ambiguity, the design used either deoxyinosines or most frequently used codons based on the codon usage in Arabidopsis genes. When used as a hybridization probe, the 135-mer hybridized to all plaques that also hybridized to cDNAs encoding Arabidopsis microsomal and plastid delta-15 desaturases (SEQ ID NOS:1 and 4). In addition, it also hybridized to plaques that did not hybridize to SEQ ID NOS:1 and 4). The latter were purified and excised as described previously. Nucleotide sequencing of the cDNA inserts in the resultant plasmids revealed DNA sequences that did not show any relatedness to any desaturase.

For another example, in the polymerase chain reaction (Innis, et al., Eds, (1990) PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego), two short pieces of the present fragment of the invention can be used to amplify a longer glycerolipid desaturase DNA fragment from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleotide sequences with one primer based on the fragment of the invention and the other on either the poly A^+ tail or a vector sequence. These oligomers may

be unique sequences or degenerate sequences derived from the nucleic acid fragments of the invention. The longer piece of homologous glycerolipid desaturase DNA generated by this method could then be used as a probe for isolating related glycerolipid desaturase genes or cDNAs from Arabidopsis or other species. The design of oligomers, including long oligomers using deoxyinosine, and "guessmers" for hybridization or for the polymerase chain reaction are known to one skilled in the art and discussed in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press). Stretches of conserved amino acids between delta-15 desaturase and other desaturases, especially desA, allow for the design of such oligomers. For example, conserved stretches of amino acids between desA and delta-15 desaturase, discussed above, are useful in designing long oligomers for hybridization as well as shorter ones for use as primers in the polymerase chain reaction. In this regard, the conserved amino acid stretch of amino acids 97 to 108 of SEQ ID NO:2 is particularly useful. Other conserved regions in SEQ ID NO:2 useful for this purpose are amino acids 299 to 309, amino acids 115 to 121, and amino acids 133 to 141. Amino acid stretch 133 to 141 in SEQ ID NO:2 shows especially good homology to several desaturases. For example, in this stretch, amino acids 133, 137, 138, 140 and 141 are conserved in plant delta-15 desaturases, cyanobacterial desA, yeast and mammalian microsomal stearoyl-CoA desaturases. Comparison of cyanobacterial des A and plant delta-15 desaturases revealed two particularly well conserved stretch of amino acids (amino acids 97-108 and amino acids 299-311 in SEQ ID NO:1) that can be used for PCR. The following sets of PCR primers were made to these regions:

SEQ ID NO	Length	Fold Degeneracy	AA positions in SEQ ID NO:2	AA Sequence
20	36	16	97-108 (S)	FVLGHDCGHGSF
21	36	16	97-108 (S)	FVLGHDCGHGSF
28	36	16	97-108 (S)	FVLGHDCGHGSF
29	36	16	97-108 (S)	FVLGHDCGHGSF
22	18	72	100-105 (S)	GHDCGH
23	18	72	100-105 (S)	GHDCGH
24	18	72	299-304 (AS)	HDIGTH
25	18	72	299-304 (AS)	HDIGTH
26	23	416	304-309 (AS)	HVIHHL
27	23	416	304-309 (AS)	HVIHHL
30	38	64	299-311 (AS)	HDIGTHVIHHLFP
31	38	64	299-311 (AS)	HDIGTHVIHHLFP

In one experiment, PCRs were performed using SEQ ID NOS:22 and 23 as sense primers and either SEQ ID NOS:24 and 25 or SEQ ID NOS:26 and 27 as antisense primers on poly A+ RNA purified from both Arabidopsis leaf and canola developing seeds. All PCRs resulted in PCR products of the correct size (ca. 630 bp). The PCR products from Arabidopsis and canola were purified and used as radiolabeled hybridization probes to screen the Lambda Yes Arabidopsis cDNA library, as described above. This led to the isolation of a pure phage, which was excised to give plasmid pYacp7. The cDNA insert in pYacp7 was partially sequenced. It's sequence showed that it encoded an incomplete desaturase polypeptide that was identical to another cDNA (in plasmid pFadx-2) isolated by low-stringency hybridization as described previously. The composite sequence derived from the partial sequences from the cDNA inserts in pFadx-2 and pYacp7 is shown in SEQ ID NO:16 and the polypeptide encoded by it in SEQ ID NO:17. As discussed previously, SEQ ID NO:17 is a putative plastid delta-15 desaturase. This is further supported by Southern blot analysis

using radiolabeled cDNA inserts from either pCF3, pACF2-2, or pYacp7 on Arabidopsis genomic DNA digest d with one of several enzymes. It shows that the different inserts hybridize to different restriction fragments and that only the inserts from pACF2-2 and pYacp7 show some cross-hybridization.

In another PCR experiment, PCR was performed using ca. 80 pmoles each of SEQ ID NOS:28 and 29 as sense primers and ca. 94 pmoles each of SEQ ID NOS:30 and 31 as antisense primers on poly A+ RNA purified from Arabidopsis mutant line 3707. This was performed using GeneAmp® RNA PCR Kit (Perkin Elmer Cetus) following manufacturer's protocol and using the following program: a) 1 cycle of 2 min at 95°C, b) 35 cycles of 1 min at 95°C (denaturation), 1 min at 50°C (annealing) and 1 min at 65°C (extension), and c) 1 cycle of 7 min at 65°C. The resulting PCR product, of the correct size (ca. 630 bp), was purified, radiolabeled, and used as a hybridization probe on a Southern blot of Arabidopsis genomic DNA as described above. While it hybridized to restriction fragments that also hybridized to SEQ ID NOS:1 (Arabidopsis microsomal delta-15 desaturase), 4 (Arabidopsis plastid delta-15 desaturase), and 16 (Arabidopsis plastid delta-15 desaturase), it also hybridized to novel fragments that did not hybridze to previously cloned desaturase cDNAs. However, even after several attempts, the radiolabeled PCR product did not hybridize to any novel cDNA clone when used as a probe on different Arabidopsis cDNA libraries: in all cases it hybridzed only to plaques that also hybridized to the known desaturase cDNAs. Furthermore, the PCR product was subcloned into a plasmid vector and after screening about a 100 of these, none gave rise to a clone with a novel desaturase sequence.

The isolation of other glycerolipid d saturases will become easier as more examples of glycerolipid desaturases are isolated using the fragments of the invention. Knowing the conserved amino acid sequences from diverse desaturases will also allow one to identify more and better consensus sequences. Such sequences can be used to make hybridization probes or amplification primers which will further aid in the isolation of different glycerolipid desaturases, including those from non-plant sources such as fungi, algae, and even cyanobacteria, as well as other membrane-associated desaturases from other organisms.

The function of the diverse nucleotide fragments encoding glycerolipid desaturases that can be isolated using the present invention can be identified by transforming plants with the isolated desaturase sequences, linked in sense or antisense orientation to suitable regulatory sequences required for plant expression, and observing the fatty acid phenotype of the resulting transgenic plants. Preferred target plants for the transformation are the same as the source of the isolated nucleotide fragments when the goal is to obtain inhibition of the corresponding endogenous gene by antisense inhibition or cosuppression. Preferred target plants for use in expression or overexpression of the isolated nucleic acid fragments are plants with known mutations in desaturation reactions, such as the Arabidopsis desaturase mutants, mutant flax deficient in delta-15 desaturation, or mutant sunflower deficient in delta-12 desaturation. Alternatively, the function of the isolated nucleic acid fragments can be determined similarly via transformation of other organisms, such as yeast or cyanobacteria, with chimeric genes containing the nucleic acid fragment and suitable regulatory

sequences followed by analysis of fatty acid composition and/or enzyme activity.

Overexpression of the Glycerolipid Desaturase Enzymes in Transgenic Species

The nucleic acid fragment(s) of the instant invention encoding functional glycerolipid desaturase(s), with suitable regulatory sequences, can be used to overexpress the enzyme(s) in transgenic organisms. Such recombinant DNA constructs may include either the native glycerolipid desaturase gene or a chimeric glycerolipid desaturase gene isolated from the same or a different species as the host organism. For overexpression of glycerolipid desaturase(s), it is preferable that the introduced gene be from a different species to reduce the likelihood of cosuppression. For example, overexpression of delta-15 desaturase in soybean, rapeseed, or other oil-producing species to produce altered levels of polyunsaturated fatty acids may be achieved by expressing RNA from the entire cDNA found in pCF3. Similarly, the isolated nucleic acid fragments encoding glycerolipid desaturases from Arabidopsis, rapeseed, and soybean can also be used by one skilled in the art to obtain substantially homologous full-length cDNAs, if not already obtained, as well as the corresponding genes as fragments of the invention. These, in turn, may be used to overexpress the corresponding desaturases in plants. One skilled in the art can also isolate the coding sequence(s) from the fragment(s) of the invention by using and/or creating sites for restriction endonucleases, as described in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press). For example, the fragment in SEQ ID NO:1 in plasmid pCF3 is flanked by Not I sites and can be isolated as a Not I fragment that can be introduced in

the sense orientation relative to suitable plant regulatory s quences. Alternatively, sites for Nco I (5'-CCATGG-3') or Sph I (5'-GCATGC-3') that allow precise removal of coding sequences starting with the initiating codon "ATG" may be engineered into the fragment(s) of the invention. For example, for utilizing the coding sequence of delta-15 desaturase from pCF3, an Sph I site can be engineered by substituting nucleotides at positions 44, 45, and 49 of SEQ ID NO:1 with G, C, and C, respectively.

Inhibition of Plant Target
Genes by Use of Antisense RNA

Antisense RNA has been used to inhibit plant target genes in a tissue-specific manner (see van der Krol et al., Biotechniques (1988) 6:958-976). Antisense inhibition has been shown using the entire cDNA sequence (Sheehy et al., Proc. Natl. Acad. Sci. USA (1988) 85:8805-8809) as well as a partial cDNA sequence (Cannon et al., Plant Molec. Biol. (1990) 15:39-47). There is also evidence that the 3' non-coding sequences (Ch'ng et al., Proc. Natl. Acad. Sci. USA (1989) 86:10006-10010) and fragments of 5' coding sequence, containing as few as 41 base-pairs of a 1.87 kb cDNA (Cannon et al., Plant Molec. Biol. (1990) 15:39-47), can play important roles in antisense inhibition.

The use of antisense inhibition of the glycerolipid desaturases may require isolation of the transcribed sequence for one or more target glycerolipid desaturase genes that are expressed in the target tissue of the target plant. The genes that are most highly expressed are the best targets for antisense inhibition. These genes may be identified by determining their levels of transcription by techniques, such as quantitative analysis of mRNA levels or nuclear run-off transcription, known to one skilled in the art.

For example, antisense inhibition of delta-15 desaturase in Brassica napus resulting in alter d levels of polyunsaturated fatty acids may be achieved by expressing antisense RNA from the entire or partial cDNA found in pBNSF3-2.

Inhibition of Plant Target Genes by Cosuppression

The phenomenon of cosuppression has also been used to inhibit plant target genes in a tissue-specific manner. Cosuppression of an endogenous gene using the entire cDNA sequence (Napoli et al., The Plant Cell (1990) 2:279-289; van der Krol et al., The Plant Cell (1990) 2:291-299) as well as a partial cDNA sequence (730 bp of a 1770 bp cDNA) (Smith et al., Mol. Gen. Genetics (1990) 224:477-481) are known.

The nucleic acid fragments of the instant invention encoding glycerolipid desaturases, or parts thereof, with suitable regulatory sequences, can be used to reduce the level of glycerolipid desaturases, thereby altering fatty acid composition, in transgenic plants which contain an endogenous gene substantially homologous to the introduced nucleic acid fragment. The experimental procedures necessary for this are similar to those described above for the overexpression of the glycerolipid desaturase nucleic acid fragments. For example, cosuppression of delta-15 desaturase in Brassica napus resulting in altered levels of polyunsaturated fatty acids may be achieved by expressing in the sense orientation the entire or partial seed delta-15 desaturase cDNA found in pBNSF3-2.

Selection of Hosts, Promoters and Enhancers

A preferred class of heterologous hosts for the expression of the nucleic acid fragments of the invention are eukaryotic hosts, particularly the cells of higher plants. Particularly preferred among the

higher plants are the oil-producing species, such as soybean (Glycine max), rapeseed (including Brassica napus, B. campestris), sunflower (Helianthus annus), cotton (Gossypium hirsutum), corn (Zea mays), cocoa (Theobroma cacao), safflower (Carthamus tinctorius), oil palm (Elaeis guineensis), coconut palm (Cocos nucifera), flax (Linum usitatissimum), and peanut (Arachis hypogaea).

Expression in plants will use regulatory sequences functional in such plants. The expression of foreign genes in plants is well-established (De Blaere et al., Meth. Enzymol. (1987) 153:277-291). The source of the promoter chosen to drive the expression of the fragments of the invention is not critical provided it has sufficient transcriptional activity to accomplish the invention by increasing or decreasing, respectively, the level of translatable mRNA for the glycerolipid desaturases in the desired host tissue. Preferred promoters include (a) strong constitutive plant promoters, such as those directing the 19S and 35S transcripts in cauliflower mosaic virus (Odell et al., Nature (1985) 313:810-812; Hull et al., Virology (1987) 86:482-493), and (b) tissue- or developmentally-specific promoters. Examples of tissue-specific promoters are the light-inducible promoter of the small subunit of ribulose 1,5-bis-phosphate carboxylase (if expression is desired in photosynthetic tissues), the maize zein protein promoter (Matzke et al., EMBO J. (1984) 3:1525-1532), and the chlorophyll a/B binding protein promoter (Lampa et al., Nature (1986) 316:750-752).

Particularly preferred promoters are those that allow seed-specific expression. This may be especially useful since seeds are the primary source of vegetable oils and also since seed-specific expression will avoid any potential deleterious effect in non-seed tissues.

Examples of s ed-specific promoters include, but are not limited to, the promoters of seed storage prot ins, which can represent up to 90% of total seed protein in many plants. The seed storage proteins are strictly regulated, being expressed almost exclusively in seeds in a highly tissue-specific and stage-specific manner (Higgins et al., Ann. Rev. Plant Physiol. (1984) 35:191-221; Goldberg et al., Cell (1989) 56:149-160). Moreover, different seed storage proteins may be expressed at different stages of seed development.

Expression of seed-specific genes has been studied in great detail (See reviews by Goldberg et al., Cell (1989) 56:149-160 and Higgins et al., Ann. Rev. Plant Physiol. (1984) 35:191-221). There are currently numerous examples of seed-specific expression of seed storage protein genes in transgenic dicotyledonous plants. These include genes from dicotyledonous plants for bean b-phaseolin (Sengupta-Gopalan et al., Proc. Natl. Acad. Sci. USA (1985) 82:3320-3324; Hoffman et al., Plant Mol. Biol. (1988) 11:717-729), bean lectin (Voelker et al., EMBO J. (1987) 6:3571-3577), soybean lectin (Okamuro et al., Proc. Natl. Acad. Sci. USA (1986) 83:8240-8244), soybean Kunitz trypsin inhibitor (Perez-Grau et al., Plant Cell (1989) 1:095-1109), soybean b-conglycinin (Beachy et al., EMBO J. (1985) 4:3047-3053; pea vicilin (Higgins et al., Plant Mol. Biol. (1988) 11:683-695), pea convicilin (Newbigin et al., Planta (1990) 180:461-470), pea legumin (Shirsat et al., Mol. Gen. Genetics (1989) 215:326-331); rapeseed napin (Radke et al., Theor. Appl. Genet. (1988) 75:685-694) as well as genes from monocotyledonous plants such as for maize 15 kD zein (Hoffman et al., EMBO J. (1987) 6:3213-3221), maize 18 kD oleosin (Lee at al., Proc. Natl. Acad. Sci. USA (1991) 888:6181-6185), barley b-hordein (Marris et al., Plant Mol. Biol. (1988)

10:359-366) and wheat glutenin (Colot et al., EMBO J. (1987) 6:3559-3564). Moreover, promoters of seed-specific genes operably linked to heterologous coding sequences in chimeric gene constructs also maintain their temporal and spatial expression pattern in transgenic plants. Such examples include use of Arabidopsis thaliana 2S seed storage protein gene promoter to express enkephalin peptides in Arabidopsis and B. napus seeds (Vandekerckhove et al., Bio/Technology (1989) 7:929-932), bean lectin and bean b-phaseolin promoters to express luciferase (Riggs et al., Plant Sci. (1989) 63:47-57), and wheat glutenin promoters to express chloramphenicol acetyl transferase (Colot et al., EMBO J. (1987) 6:3559-3564).

Of particular use in the expression of the nucleic acid fragment of the invention will be the heterologous promoters from several soybean seed storage protein genes such as those for the Kunitz trypsin inhibitor (Jofuku et al., Plant Cell (1989) 1:1079-1093; glycinin (Nielson et al., Plant Cell (1989) 1:313-328), and b-conglycinin (Harada et al., Plant Cell (1989) 1:415-425). Promoters of genes for a- and b-subunits of soybean β-conglycinin storage protein will be particularly useful in expressing the mRNA or the antisense RNA in the cotyledons at mid- to late-stages of seed development (Beachy et al., EMBO J. (1985) 4:3047-3053) in transgenic plants. This is because there is very little position effect on their expression in transgenic seeds, and the two promoters show different temporal regulation. The promoter for the a-subunit gene is expressed a few days before that for the b-subunit gene. This is important for transforming rapeseed where oil biosynthesis begins about a week before seed storage protein synthesis (Murphy et al., J. Plant Physiol. (1989) 135:63-69).

Also of particular use will be promoters of genes expressed during early embryogenesis and oil biosynthesis. The native regulatory sequences, including the native promoters, of the glycerolipid desaturase genes expressing the nucleic acid fragments of the invention can be used following their isolation by those skilled in the art. Heterologous promoters from other genes involved in seed oil biosynthesis, such as those for B. napus isocitrate lyase and malate synthase (Comai et al., Plant Cell (1989) 1:293-300), delta-9 desaturase from safflower (Thompson et al. Proc. Natl. Acad. Sci. USA (1991) 88:2578-2582) and castor (Shanklin et al., Proc. Natl. Acad. Sci. USA (1991) 88:2510-2514), acyl carrier protein (ACP) from Arabidopsis (Post-Beittenmiller et al., Nucl. Acids Res. (1989) 17:1777), B. napus (Safford et al., Eur. J. Biochem. (1988) 174:287-295),and B. campestris (Rose et al., Nucl. Acids Res. (1987) 15:7197), b-ketoacyl-ACP synthetase from barley (Siggaard-Andersen et al., Proc. Natl. Acad. Sci. USA (1991) 88:4114-4118), and oleosin from Zea mays (Lee et al., Proc. Natl. Acad. Sci. USA (1991) 88:6181-6185), soybean (Genbank Accession No: X60773) and B. napus (Lee et al., Plant Physiol. (1991) 96:1395-1397) will be of use. If the sequence of the corresponding genes is not disclosed or their promoter region is not identified, one skilled in the art can use the published sequence to isolate the corresponding gene and a fragment thereof containing the promoter. The partial protein sequences for the relatively-abundant enoyl-ACP reductase and acetyl-CoA carboxylase are also published (Slabas et al., Biochim. Biophys. Acta (1987) 877:271-280; Cottingham et al., Biochim. Biophys. Acta (1988) 954:201-207) and one skilled in the art can use these sequences to isolate the corresponding seed genes with their promoters. Similarly, the fragments of the

present invention encoding glycerolipid desaturases can be used to obtain promoter regions of the corresponding genes for use in expressing chimeric genes.

Attaining the proper level of expression of the nucleic acid fragments of the invention may require the use of different chimeric genes utilizing different promoters. Such chimeric genes can be transferred into host plants either together in a single expression vector or sequentially using more than one vector.

It is envisioned that the introduction of enhancers or enhancer-like elements into the promoter regions of either the native or chimeric nucleic acid fragments of the invention will result in increased expression to accomplish the invention. This would include viral enhancers such as that found in the 35S promoter (Odell et al., Plant Mol. Biol. (1988) 10:263-272), enhancers from the opine genes (Fromm et al., Plant Cell (1989) 1:977-984), or enhancers from any other source that result in increased transcription when placed into a promoter operably linked to the nucleic acid fragment of the invention.

Of particular importance is the DNA sequence element isolated from the gene for the a-subunit of b-conglycinin that can confer 40-fold seed-specific enhancement to a constitutive promoter (Chen et al., Dev. Genet. (1989) 10:112-122). One skilled in the art can readily isolate this element and insert it within the promoter region of any gene in order to obtain seed-specific enhanced expression with the promoter in transgenic plants. Insertion of such an element in any seed-specific gene that is expressed at different times than the b-conglycinin gene will result in expression in transgenic plants for a longer period during seed development.

The invention can also be accomplished by a variety of other methods to obtain the desired end. In one form, the invention is based on modifying plants to produce increased levels of glycerolipid desaturases by virtue of introducing more than one copy of the foreign gene containing the nucleic acid fragments of the invention. In some cases, the desired level of polyunsaturated fatty acids may require introduction of foreign genes for more than one kind of glycerolipid desaturase.

Any 3' non-coding region capable of providing a polyadenylation signal and other regulatory sequences that may be required for the proper expression of the nucleic acid fragments of the invention can be used to accomplish the invention. This would include 3' ends of the native glycerolipid desaturase(s), viral genes such as from the 35S or the 19S cauliflower mosaic virus transcripts, from the opine synthesis genes, ribulose 1,5-bisphosphate carboxylase, or chlorophyll a/b binding protein. There are numerous examples in the art that teach the usefulness of different 3' non-coding regions.

Transformation Methods

Various methods of transforming cells of higher plants according to the present invention are available to those skilled in the art (see EPO Pub. 0 295 959 A2 and 0 318 341 A1). Such methods include those based on transformation vectors utilizing the Ti and Ri plasmids of *Agrobacterium spp*. It is particularly preferred to use the binary type of these vectors. Ti-derived vectors transform a wide variety of higher plants, including monocotyledonous and dicotyledonous plants (Sukhapinda et al., Plant Mol. Biol. (1987) 8:209-216; Potrykus, Mol. Gen. Genet. (1985) 199:183). Other transformation methods are available to those skilled in the art, such as direct uptake of foreign DNA constructs

(see EPO Pub. 0 295 959 A2), techniques of electroporation (Fromm et al., Nature (1986) (London) 319:791) or high-velocity ballistic bombardment with metal particles coated with the nucleic acid constructs (Kline et al., Nature (1987) (London) 327:70). Once transformed, the cells can be regenerated by those skilled in the art.

Of particular relevance are the recently described methods to transform foreign genes into commercially important crops, such as rapeseed (De Block et al., Plant Physiol. (1989) 91:694-701), sunflower (Everett et al., Bio/Technology (1987) 5:1201), and soybean (Christou et al., Proc. Natl. Acad. Sci USA (1989) 86:7500-7504.

Application to RFLP Technology

The use of restriction fragment length polymorphism (RFLP) markers in plant breeding has been well-documented in the art (Tanksley et al., Bio/Technology (1989) 7:257-264). The nucleic acid fragments of the invention can be used as RFLP markers for traits linked to expression of glycerolipid desaturases. These traits will include altered levels of unsaturated fatty acids. The nucleic acid fragment of the invention can also be used to isolate the glycerolipid desaturase gene from variant (including mutant) plants with altered levels of unsaturated fatty acids. Sequencing of these genes will reveal nucleotide differences from the normal gene that cause the variation. Short oligonucleotides designed around these differences may be used as hybridization probes to follow the variation in polyunsaturates. Oligonucleotides based on differences that are linked to the variation may be used as molecular markers in breeding these variant oil traits.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. All publications, including patents and non-patent literature, referred to in this specification are expressly incorporated by reference herein.

EXAMPLE 1

ISOLATION OF GENOMIC DNA FLANKING THE T-DNA SITE OF INSERTION IN ARABIDOPSIS THALIANA MUTANT LINE 3707

Identification of an *Arabidopsis thaliana* T-DNA Mutant with Low Linolenic Acid Content

A population of *Arabidopsis thaliana* (geographic race Wassilewskija) transformants containing the T-DNA of *Agrobacterium tumefaciens* was generated by seed transformation as described by Feldmann et al., (Mol. Gen. Genetics (1987) 208:1-9). In this population the transformants contain DNA sequences encoding the pBR322 bacterial vector, nopaline synthase, neomycin phosphotransferase (NPTII, confers kanamycin resistance), and b-lactamase (confers ampicillin resistance) within the T-DNA border sequences. The integration of the T-DNA into different areas of the chromosomes of individual transformants may cause a disruption of plant gene function at or near the site of insertion, and phenotypes associated with this loss of

gene function can be analyz d by screening the population for the phenotype.

T3 seed was generated from the wild type seed treated with Agrobacterium tumefaciens by two rounds of self-fertilization as described by Feldmann et al., (Science (1989) 243:1351-1354). These progeny were segregating for the T-DNA insertion, and thus for any mutation resulting from the insertion. Approximately 100 seeds of each of 6000 lines were combined and the fatty acid content of each of the 6000 pooled samples was determined by gas chromatography of the fatty acyl methyl esters essentially as described by Browse et al., (Anal. Biochem. (1986) 152:141-145) except that 2.5% H_2SO_4 in methanol was used as the methylation reagent and samples were heated for 1.5 h at 80°C to effect the methanolysis of the seed triglycerides. A line designated "3707" produced seeds that gave an altered fatty acid profile compared to that of the total population. T3 plants were grown from individual T3 seeds produced by line 3707 and self-fertilized to produce T4 seeds on individual plants that were either homozygous wild type, homozygous mutant, or heterozygous for the mutation. The percent fatty acid compositions of a representative subsample of the entire population, of the pooled 3707 T3 seeds, and of a homozygous T4 mutant segregant are shown in Table 4.

TABLE 4

Fatty Acid Methyl Ester	T3 Pools from lines 3501-4000 average and (std. deviation)	3707 T3 Pool	3707 Homozygous T4 Segregant
palmitic	7.4 (0.37)	7.0	6.4
stearic	3.0 (0.22)	2.9	3.0
oleic	17.0 (1.5)	17.7	15.9
linoleic	29.3 (0.78)	35.0	42.4
linolenic	16.1 (1.1)	10.2	3.1
eicosenoic	20.2 (0.73)	20.5	23.6

The phenotype of the segregating T3 pool of line 3707 (high linoleic acid, low linolenic acid) was intermediate between that of the population subsample and the homozygous T4 mutant seeds suggesting that line 3707 harbored a mutation at a locus which controls the conversion of linoleic to linolenic acid in the seed. Still, it was not apparent whether the mutant phenotype in line 3707 was the result of a T-DNA insertion. Therefore, Applicants checked a segregating T4 population to determine whether the mutant fatty acid phenotype cosegregated with the nopaline synthase activity and kanamycin resistance encoded by the T-DNA insert. A total of 263 T4 plants were grown and assayed for the presence of nopaline in leaf extracts (Errampalli et al., The Plant Cell (1991) 3:149-157). In addition, T5 seeds were collected from each of the T4 plants and samples of 10-50 seeds were taken to determine the seed fatty acid composition and to determine their ability to germinate in the presence of kanamycin (Feldmann, et al., (1989) Science 243:1351-1354). The 263 plants fell into 3 classes as in Table 5.

TABLE 5

Number of Individuals	Phenotype
63	T4 plants: little or no nopaline present; T5 seeds: wild type fatty acid composition, all kanamycin sensitive
134	T4 plants: nopaline present; T5 seeds: heterozygous fatty acid composition similar to 3707 T3 pool, segregating for kanamycin resistance
64	T4 plants: nopaline present; T5 seeds homozygous mutant fatty acid composition, all kanamycin resistant

The cosegregation of the fatty acid phenotype with the phenotypes conferred by T-DNA sequences in an approximately 1:2:1 pattern provided strong evidence that the mutation in line 3707 was the result of a T-DNA insertion. Further experiments were then conducted with the intent of using probes containing T-DNA sequences to clone the T-DNA insert and flanking genomic DNA from line 3707.

Preparation of Genomic DNA from Homozygous 3707 Plants

Seeds from a homozygous line derived from *Arabidopsis thaliana* (geographic race Wassilewskija (WS)) line 3707 were surface sterilized for 5 min at room temperature in a solution of 5.25% sodium hypochlorite (w/v)/0.15% Tween 20 (v/v), then washed several times in sterile distilled water, with a final rinse in 50% ethanol. Immediately following the ethanol wash, the seeds were transferred to sterile filter paper to dry. One to three seeds were then transferred to 250-mL flasks containing 50 mL of sterile Gamborgs B5 media (Gibco, 500-1153EA), pH 6.0. Cultures were incubated at 22°C, 70 $\mu E \cdot /m^{-2} \cdot sec^{-1}$ of continuous light for approximately three weeks, after which time the root tissue was harvested, made into 10 g aliquots (wet weight), lyophilized, and stored at -20°C.

Using a variation of the procedure of Shure et al., (Cell (1983) 35:225-233) genomic DNA was isolated from the root tissue. Two aliquots of lyophilized tissue were ground to a fine powder using a mortar and pestle. The ground tissue was added to a flask containing 85 mL of lysis buffer (7 M urea, 0.35 M NaCl, 0.05 M Tris-HCl, pH 8.0, 0.02 M EDTA, 1% Sarkosyl, 5% phenol) and mixed gently with a glass rod to obtain a homogeneous suspension. To this suspension an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) (equilibrated with 10 mM Tris, pH 8, 1 mM EDTA) was added. After the addition of 8.5 mL of 10% SDS the mixture was swirled on a rotating platform for 15 min at room temperature. After centrifugation at 2000xg for 15 min, the upper aqueous phase was removed to a new tube and extracted two more times, as above, but without the addition of SDS. To the final aqueous phase was added 1/20th the volume of 3 M potassium acetate, pH 5.5 and two times the volume of ice cold 100% ethanol. Precipitation of the DNA was facilitated by incubation at -20°C for one hour followed by centrifugation at 12,000xg for 10 min. The resulting pellet was resuspended in 3 mL of 10 mM Tris, pH 8, 1 mM EDTA to which was added 0.95 g of cesium chloride (CsCl) and 21.4 µL of 10 mg/mL ethidium bromide (EtBr) per mL of solution. The DNA was then purified by centrifugation to equilibrium in a CsCl/EtBr density gradient for 16 h at 15°C, 265,000xg. After removal from the gradient, the DNA was extracted with isopropanol saturated with TE buffer (10 mM Tris, pH 8; 1 mM EDTA) and CsCl to remove EtBr and then dialyzed overnight at 4°C against 10 mM Tris, pH 8, 1 mM EDTA to remove CsCl. The DNA was removed from dialysis and the concentration was determined using the Hoechst fluorometric assay in which an aliquot of DNA is added to 3 mL of 1.5×10^{-6} M bis-

benzimide (Hoechst 33258, Siga) in 1X SSC (0.15 M NaCl, 0.015 M sodium citrate), pH 7.0, incubated at room temperature for 5 min, and read on a fluorometer at excitation 360, emission 450, against a known set of DNA standards.

Plasmid Rescue and Analysis

Five micrograms of genomic DNA from the homozygous 3707 mutant, prepared as described above, was digested with 20 units of either Bam HI or Sal I restriction enzyme (Bethesda Research Laboratory) in a 50 µL reaction volume according to the manufacturer's specifications. After digestion the DNA was extracted with buffer-saturated phenol (Bethesda Research Laboratory) followed by precipitation in ethanol. The resulting pellet was resuspended in a final volume of 10 µL of 10 mM Tris, pH 8, and the concentration of the DNA was determined using the Hoechst fluorometric assay as above.

To facilitate circularization, as opposed to end-to-end joining, a dilute ligation reaction was set up containing 250 ng of Bam HI or Sal I digested genomic DNA, 3 Weiss units of T4 DNA ligase (Promega), 50 µL of 10X ligase buffer (30 mM Tris-HCl, pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP) and 5 µL of 100 mM ATP in a 500 µL reaction volume. The reaction was incubated for 16 h at 16°C, heated for 10 min at 70°C, and extracted once with buffer saturated phenol (Bethesda Research Laboratory). The DNA was then precipitated with the addition of two volumes of 100% ethanol and 1/10th volume of 7.5 M ammonium acetate. The resulting pellet was resuspended in a final volume of 10 µL of 10 mM Tris, pH 8, and the concentration of the DNA was determined using the Hoechst fluorometric assay as above.

Competent DH10B cells (Bethesda Research Laboratory) were transfected with 50 ng of ligated DNA

at a concentration of 10 ng of DNA per 100 µL of cells according to the manufacturer's specifications. Transformants from Sal I or Bam HI digests were selected on LB plates (10 g Bacto-tryptone, 5 g Bacto-yeast extract, 5 g NaCl, 15 g agar per liter, pH 7.4) containing 100 µg/mL ampicillin or 25 µg/mL kanamycin sulfate, respectively. Ampicillin-resistant (Amp^r; ampicillin sensitivity, Amp^s) Sal I tranformants were screened for the presence of the kanamycin resistance (Kan^r; kanamycin sensitivity, Kan^s) gene by picking primary tranformants and stabbing them first to LB plates containing 100 µg/mL ampicillin then to LB plates containing 25 µg/mL kanamycin. After overnight incubation at 37°C the plates were scored for Amp^r/Kan^s colonies. Kanamycin-resistant Bam HI transformants were screened for the presence of the ampicillin resistance gene by picking primary transformants and stabbing them first to LB plates containing 25 µg/mL kanamycin and then to LB plates containing 100 µg/mL ampicillin. After overnight incubation at 37°C the plates were scored for Kan^r/Amp^r colonies.

Cultures were made of 192 Amp^r/Kan^s Sal I transformants and 85 Kan^r/Amp^r Bam HI transformants directly into deep-well microtiter plates containing 200 µL of LB broth (10 g Bacto-tryptone, 5 g Bacto-yeast extract, 5 g NaCl per liter) with 100 µg/mL ampicillin. Using the Schleicher and Schuell Minifold I apparatus and Nytran membranes, dot blots were set up, in duplicate, using the following conditions: 50 µL of culture was diluted into 150 µL of 5X SSC, the culture was lysed and the DNA denatured by the addition of 150 µL of 0.5 M NaOH, 1.5 M NaCl solution for 3 min at room temperature, the filter was removed from the apparatus and neutralized in 0.5 M Tris, pH 8, 1.5 M NaCl, the DNA was then UV cross-linked to the filters

using the Stratagene Stratalinker, and the filters were heated for 2 h at 80°C and stored at room temperature.

To determine whether T-DNA was contained within any of the rescued plasmids, the dot blots were probed with portions of the right and left borders of T-DNA. The right border probe consisted of a 2.2 kb Hind III-Dra I fragment of DNA obtained from plasmid H23pKC7 (composed of the 3.2 kb Hind III 23 fragment from Ti plasmid pTiC58 (Lemmers et al., J. Mol. Biol. (1989) 144;353-376) cloned into plasmid vector pKC7 (Maniatis et al., Molecular Cloning, A Laboratory Manual (1982) Cold Spring Harbor Laboratory Press)), and the left border probe consisted of a 2.9 kb Hind III-Eco RI fragment obtained from plasmid H10pKC7 (composed of the 6.5 kb Hind III 10 fragment from Ti plasmid pTiC58 (Lemmers et al., J. Mol. Biol. (1989) 144:353-376) cloned into plasmid vector pKC7 (Maniatis et al., Molecular Cloning, A Laboratory Manual (1982) Cold Spring Harbor Laboratory Press)) using standard digestion, electrophoresis, and electroelution conditions as described in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed (1989) Cold Spring Harbor Laboratory Press). Final DNA purification was obtained by passage of the eluted DNA over an Elutip-D column (Schleicher and Schuell) using the manufacturer's specifications. Concentration of the DNA was determined using the Hoechst fluorometric assay as above. Approximately 100 ng of each probe was labeled with a[^{32}P]dCTP using a Random Priming Kit from Bethesda Research Laboratories under conditions recommended by the manufacturer. Labeled probe was separated from unincorporated a[^{32}P]dCTP by passing the reaction through a Sephadex G-25 spun column under standard conditions as described in Sambrook et al., (Molecular

Cloning, A Laboratory Manual, 2nd ed. (1989) Cold Spring Harbor Laboratory Press).

The filters were pre-hybridized in 150 mL of buffer consisting of 6X SSC, 10X Denhardt's solution, 1% SDS, and 100 µg/mL denatured calf thymus DNA for 16 h at 42°C. The denatured, purified, labeled probe was added to the pre-hybridized filters following transfer of the filters to 50 mL of hybridization buffer consisting of 6X SSC, 1% SDS, 10% dextran sulfate, and 50 µg/mL denatured calf thymus DNA. Following incubation of the filters in the presence of the probe for 16 h at 65°C, the filters were washed twice in 150 mL of 6X SSC, 0.5% SDS, twice in 1X SSC, 1% SDS and once in 0.1X SSC, 1% SDS, all at 65°C. The washed filters were subjected to autoradiography on Kodak XAR-2 film at 80°C overnight.

Of the 85 Bam HI candidates, 63 hybridized with the left border probe and none hybridized with the right border probe. Of the 192 Sal I candidates, 31 hybridized with the left border probe, 4 hybridized with the right border probe, and none hybridized with both probes. Twelve of the Bam HI candidates, 7 positive and 5 negative for the presence of the left border of T-DNA, were further analyzed by restriction digests.

DNA from the Bam HI candidates was made by the alkaline lysis miniprep procedure of Birmbiom et al., (Nuc. Acid Res. (1979) 7:1513-1523), as described in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press). The plasmid DNA was digested with Eco RI restriction enzyme (Bethesda Research Laboratories) in accordance with the manufacturer's specifications and electrophoresed through a 0.8% agarose gel in 1X TBE buffer (0.089 M Tris-borate, 0.089 M boric acid, 0.002 M EDTA). All of the Bam HI candidates which hybridized with the left border probe of T-DNA had the same Eco RI

restriction pattern, which indicated the presence of 14.2 kb of T-DNA and 1.4 kb of putative plant genomic DNA in these clones.

DNA from Sal I candidates was isolated, restriction-analyzed using Eco RI, Bam HI and Sal I enzymes, and electrophoresed through a 0.8% agarose gel, as above. All of the Sal I candidates which hybridized with the left border probe of T-DNA included 2.9 kb of putative plant DNA. Contained within this 2.9 kb fragment was a 1.4 kb Bam HI-Eco RI fragment as seen with the Bam HI rescued plasmids, suggesting that the 1.4 kb fragment was a subset of the 2.9 kb fragment and that it was adjacent to the left border of the T-DNA at its site of insertion into the plant genome. Sequence analysis of one Sal I candidate (pS1) using a primer homologous to the left border sequence of T-DNA, revealed that the sequence of pS1 was colinear with the sequence of the T-DNA left border (Yadav et al., Proc. Natl. Acad. Sci. USA (1982) 79:6322-6326) up to nucleotide 65, followed by non-T-DNA (putative plant) sequences.

Southern Analysis with Putative Plant
<u>DNA from Rescued Plasmids</u>

DNA from the seven Bam HI candidates which hybridized with the left border of the T-DNA was pooled and a portion was digested with Eco RI and Bam HI restriction endonucleases and electrophetically separated on a 0.8% agarose gel in 1X TBE buffer. After excising a 1.4 kb Eco RI-Bam HI fragment from the agarose gel, the 1.4 kb fragment was purified by use of a Gene Clean Kit from Bio 101. Fifty nanograms of the resulting DNA fragment was labeled with a[^{32}P]dCTP using a Random Priming Kit (Bethesda Research Laboratory) under conditions recommended by the manufacturer.

Three micrograms of total genomic DNA from homozygous wild-type *Arabidopsis* and homozygous 3707 mutant *Arabidopsis* plants was digested to completion with one of the following restriction enzymes: Sal I, Hind III, Eco RI, Cla I, and Bam HI under conditions suggested by the manufacturer. The digested DNA was subjected to electrophoresis and Southern transfer to Hybond-N membranes (Amersham) as described in Sambrook et al. (Molecular Cloning, A Laboratory Approach, 2nd. ed. (1989) Cold Spring Harbor Laboratory Press). After Southern transfer, the membranes were exposed to UV light using the Stratalinker (Stratagene) as per the manufacturer's instructions, air dried, and heated at 68°C for 2 h.

The filters were prehybridized in 1 M NaCl, 50 mM Tris-Cl, pH 7.5, 1% sodium dodecyl sulfate, 5% dextran sulfate, 100 µg/mL of denatured salmon sperm DNA at 65°C overnight. Fifty nanograms of the radiolabeled 1.4 kb Eco RI-Bam HI plant DNA fragment prepared above was added to the prehybridization solution containing the Southern blot and further incubated at 65°C overnight. The filter was washed for 10 min twice in 200 mL 2X SSPE, 0.1% sodium dodecyl sulfate at 65°C and for 10 min in 200 mL 0.5% SSPE, 0.1% sodium dodecyl sulfate at 65°C. Hybridizing fragments were detected by autoradiography. The analysis confirmed that the probe fragment contained plant DNA and that the T-DNA integration site was in a 2.8 kb Bam HI, a 5.2 kb Hind III, a 3.5 kb Sal I, a 5.5 kb Eco RI, and an approximately 9 kb Cla I fragment of wild type *Arabidopsis* DNA.

Isolation of Lambda Clones Containing the Wild Type *Arabidopsis Delta-15 Desaturase Gene*

The 1.4 kb Eco RI-Bam HI fragment (see above) was used as a probe to screen a lGem-11 library made from

genomic DNA isolated from wildtype Arabidopsis thaliana plants, geographic race WS. To construct the library, genomic DNA was partially digested with Sau3A enzyme, and size-fractionated over a salt gradient as described in Sambrook et al. (Molecular Cloning, A Laboratory Approach, 2nd ed. (1989) Cold Spring Harbor Laboratory Press). The size-fractionated DNA was then cloned into Bam HI-digested lGem-11 phage DNA (Promega) following the protocol outlined by the manufacturer. About 25,000 plaque-forming units of phage each were plated on five 150 mm petri plates containing a lawn of KW251 cells on NZY agar media (5 g NaCl, 2 g $MgSO_4 \cdot 7H_2O$, 5 g yeast extract, 10 g NZ Amine (casein hydrolysate from ICN Pharmaceuticals), 15 g agar per liter; pH 7.5). The plaques were adsorbed onto nylon membranes (Colony/Plaque Screen, New England Nuclear), in duplicate, and prepared according to the manufacturer's instructions with the addition of a 2 h incubation at 80°C after air drying the filters. The filters were prehybridized at 65°C in hybridization buffer (1% BSA, 0.5 M NaP_i, pH 7.2, (NaH_2PO_4 and Na_2HPO_4), 10 mM EDTA, and 7% SDS) for 4 h, after which time they were transferred to fresh buffer containing the denatured radiolabeled probe (see above) and incubated overnight at 65°C. The filters were rinsed twice with 0.1X SSC, 1% SDS at 65°C for 30 min each and subjected to autoradiography on Kodak XA-R film at 80°C overnight. Seven positively-hybridizing plaques were subjected to plaque purification as described in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press).

Small scale (5 mL) liquid lysates from each of the 7 clones were prepared and titered on KW251 bacteria as described in Sambrook et al. (Molecular Cloning, A Laboratory Manual, 2nd ed (1989), Cold Spring Harbor

Laboratory Press). Phage DNA was isolated using a variation of the method of Chisholm (Biotechniques (1989) 7:21-23) in which the initial lysate was made according to Sambrook et al. (Molecular Cloning, A Laboratory Manual, 2nd ed (1989), Cold Spring Harbor Laboratory Press) the concentration of DNase I and RNase I (Sigma) was reduced by half, and the PEG precipitation step was increased to 16 h. Based on restriction analysis using Hind III, Sal I and Xho I enzymes, the original 7 positive phage fell into 5 different classes. While the average insert size was approximately 15 kb, taken together the clones spanned a 40 kb region of genomic DNA. Through restriction mapping using 4 different enzymes (Hind III, Bam HI, Kpn I, and Sal I) singly, and in pair-wise combinations, accompanied by Southern analysis with the 1.4 kb Eco RI-Bam HI probe (as above) and other probes obtained from the 1 clones themselves, a partial map was obtained in which all 5 clones (11111, 141A1, 14211, 14311 and 14411) were found to share an approximately 3 kb region of homology near the site of T-DNA insertion. Via restriction and Southern analysis, Applicants ascertained that a 5.2 kb Hind III fragment present in clones 1111, 41A1, and 4411 also spanned the site of the T-DNA insertion. This fragment was excised from lambda clone 41A1, inserted into the Hind III site of the pBluescript vector (Stratagene), and the resulting plasmid, designated pF1, was prepared and isolated using standard protocols. This Hind III fragment was subsequently used to probe an <u>Arabidopsis</u> cDNA library (see below).

EXAMPLE 2

CLONING OF *ARABIDOPSIS THALIANA* DELTA-15 DESATURASE cDNA USING GENOMIC DNA FLANKING THE T-DNA SITE OF INSERTION IN *ARABIDOPSIS THALIANA* MUTANT LINE 3707 AS A HYBRIDIZATION PROBE

The 5.2 kb Hind III fragment from plasmid pF1 was purified by electrophoresis in agarose after digestion of the plasmid with Hind III and radiolabeled with ^{32}P as described above. For the preparation of an *Arabidopsis* cDNA library, polyadenylated mRNA was prepared from 3 day-old, etiolated *Arabidopsis* (ecotype Columbia) seedling hypocotyls using standard protocols (Sambrook, et al., Molecular Cloning: A Laboratory Manual, 2nd Ed. (1989) Cold Spring Harbor Laboratory Press). Five micrograms of this mRNA were used as template with an oligo d(T) primer, and Moloney Murine Leukemia Virus reverse transcriptase (Pharmacia) was used to catalyze first strand cDNA synthesis. Second-strand cDNA was made according to Gubler et al., (Gene (1983) 25:263-272) except that DNA ligase was omitted. After the second strand synthesis, the ends of the cDNA were made blunt by reaction with the Klenow fragment of DNA polymerase and ligated to Eco RI/Not I adaptors (Pharmacia). The cDNA's were purified by spun-column chromatography using Sephacryl S-300 and size-fractionated on a 1% low melting point agarose gel. Size-selected cDNAs (1-3 kb) were removed from the gel using agarase (New England Biolabs) and purified by phenol:chloroform extraction and ethanol precipitation. One hundred nanograms of the cDNA was co-precipitated with 1 µg of 1 ZAP II (Stratagene) Eco RI-digested, dephosphorylated arms. The DNAs were ligated in a volume of 4 µL overnight, and the ligation mix was packaged *in vitro* using the Gigapack II Gold packaging extract (Stratagene).

Approximately 80,000 phage were screen d for positively hybridizing plaques using the radiolab led 5.2 kb Hind III fragment as a probe essentially as described above and in Sambrook et al., (Molecular Cloning: A Laboratory Manual, 2nd ed. (1989) Cold Spring Harbor Laboratory Press). Replica filters of the phage plaques were soaked in 1 M NaCl, 50 mM Tris-HCl, pH 7.5, 1% SDS, 5% dextran sulfate, 0.1 mg/mL denatured salmon sperm DNA during the pre-hybridization step (8 hr at 65°C) and then probe was added and the hybridization proceeded over 16 hr at the same temperature. Filters were washed sequentially with 2X SSPE, 0.1% SDS at room temperature for 5 min and then again with fresh solution for 10 min, and finally with 0.5X SSPE, 0.1% SDS at 65°C for 5 min. Approximately 20 positively hybridizing plaques were identified in the primary screen. Four of these were picked and subjected to two further rounds of screening and purification. From the tertiary screen, four pure phage plaques were isolated. Plasmid clones containing the cDNA inserts were obtained through the use of a helper phage according to the *in vivo* excision protocol provided by Stratagene. Double-stranded DNA was prepared using the alkaline lysis method as previously described, and the resulting plasmids were size-analyzed by electrophoresis in agarose gels. The largest one of these, designated pCF3, contained an approximately 1.4 kb insert which was sequenced using Sequenase T7 DNA polymerase (US Biochemical Corp.) and the manufacturer's instructions, beginning with primers homologous to vector sequences that flank the cDNA insert and continuing serially with primers designed from the newly acquired sequences as the sequencing experiment progressed. The sequence of this insert is shown in SEQ ID NO:1.

EXAMPLE 3

CLONING OF AN ARABIDOPSIS CDNA ENCODING A PLASTID DELTA-15 FATTY ACID DESATURASE

A related fatty acid desaturase was cloned in a similar fashion, except that the probe used was not derived from a PCR reaction on pCF3, but rather was the actual 1.4 kb Not I fragment isolated from pCF3 which was purified and radiolabeled as described above.

Approximately 80,000 phage from the *Arabidopsis* etiolated hypocotyl cDNA library described above were plated out and screened essentially as before, except as indicated below. The filters were soaked in 1 M NaCl, 50 mM Tris-HCl, pH 7.5, 1% SDS, 5% dextran sulfate, 0.1 mg/mL denatured salmon sperm DNA during the pre-hybridization step (8 hr at 50°C). Then probe was added and the hybridization proceeded over 16 hr at the same temperature. Filters were washed sequentially with 2X SSPE, 0.1% SDS at room temperature for 5 min and then again with fresh solution for 10 min, and finally with 0.5X SSPE, 0.1% SDS at 50°C for 5 min. Approximately 17 strongly hybridizing and 17 weakly hybridizing plaques were identified in the primary screen. Four of the weakly hybridizing plaques were picked and subjected to one to two further rounds of screening with the radiolabeled probe as above until they were pure. To ensure that these were not delta-15 desaturase clones, they were further analyzed to determine whether they hybridized to a delta-15 desaturase 3' end-specific probe. The probe used was an 18 bp oligonucleotide which is complementary in sequence (i.e., antisense) to nucleotides 1229 - 1246 of SEQ ID NO:1. The probe was radiolabeled with gamma-^{32}P ATP using T4 polynucleotide kinase and hybridized to filters containing DNA from the isolated clones in 6X SSC, 5X Denhardt's, 0.1 mg/mL denatured salmon sperm DNA, 1 mM EDTA, 1% SDS at 44°C

overnight. The filters were washed twice in 6X SSC, 0.1% SDS for 5 min at room temperature, then in 6X SSC, 0.1% SDS at 44°C for 3-5 min. After autoradiography of the filters, one of the clones failed to show hybridization to this probe. This clone was picked, and a plasmid clone containing the cDNA insert was obtained through the use of a helper phage according to the in vivo excision protocol provided by Stratagene. Double-stranded DNA was prepared using the alkaline lysis method as previously described, and the resulting plasmid was size-analyzed by electrophoresis in agarose gels following either Not I digestion or digestion with both Nco I and Bgl II. The results were consistent with the presence in this plasmid, designated pCM2, of an approximately 1.3 kb cDNA insert which lacked a 0.7 kb Nco I - Bgl II fragment characteristic of the Arabidopsis delta-15 desaturase cDNA of pCF3. (This fragment corresponds to the DNA located between the Nco I site at nucleotides 474-479 and the Bgl II site at nucleotides 1164-1169 in SEQ ID NO:1). The complete nucleotide sequence of pCM2 is shown in SEQ ID NO:4.

EXAMPLE 4

CLONING OF PLANT FATTY ACID DESATURASE cDNAs FROM OTHER SPECIES BY HYBRIDIZATION TECHNIQUES

An approximately 1.4 kb fragment containing the Arabidopsis delta-15 desaturase coding sequence of SEQ ID NO:1 was obtained from plasmid pCF3 through the use of the polymerase chain reaction (PCR). Primers (M13(-20) and T7-17mer primers, 1991 Stratagene Catalogue numbers 300303 and 300302, respectively) flanking the pCF3 insert were used in the PCR which was carried out essentially as described in the instructions provided by the vendor in the Perkin-Elmer/Cetus PCR kit. This fragment was digested with Not I to remove vector sequences, purified by agarose gel electro-

phoresis, and radiolabeled with ^{32}P as previously described.

EXAMPLE 5

CLONING OF BRASSICA NAPUS SEED cDNAs ENCODING DELTA-15 FATTY ACID DESATURASES

A cDNA library from developing Brassica napus seeds was constructed using the polyadenylated mRNA fraction contained in a polysomal RNA preparation from developing Brassica napus seeds. Polysomal RNA was isolated following the procedure of Kamalay et al., (Cell (1980) 19:935-946) from seeds 20-21 days after pollination. The polyadenylated mRNA fraction was obtained by affinity chromatography on oligo-dT cellulose (Aviv et al., Proc. Natl. Acad. Sci. USA (1972) 69:1408-1411). Four micrograms of polyadenylated mRNA were reverse transcribed and used to construct a cDNA library in lambda phage (Uni-ZAP™ XR vector) using the protocol described in the ZAP-cDNA™ Synthesis Kit (1991 Stratagene Catalog, Item # 200400).

For the purpose of cloning the Brassica napus seed cDNAs encoding delta-15 fatty acid desaturases, the Brassica napus seed cDNA library was screened several times using the inserts from the Arabidopsis cDNAs pCF3 and pCM2 as radiolabelled hybridization probes. One of the Brasssica napus cDNAs obtained in these screens was used as hybridization probe in a subsequent screen.

For each screening experiment approximately 300,000 phages were screened under low stringency hybridization conditions. The filter hybridizations were carried out in 50 mM Tris pH 7.6, 6X SSC, 5X Denhardt's, 0.5% SDS, 100 ug denatured calf thymus DNA at 50°C overnight and the p[ost hybridization washes were performed in 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then

repeated twic with 0.2X SSC, 0.5% SDS at 50°C for 30 min.

Using the Arabidopsis cDNA insert of pCM2 as a probe in a low stringency screen five strongly hybridizing phages were identified. These phages were purified and excised according to the protocols described in the ZAP-cDNA™ Synthesis Kit and pBluescript II Phagemid Kit (1991 Stratagene Catalog, Item # 200400 and 212205). One of these, designated pBNSF3-f2, contained a 1.3 kb insert. pBNSF3-f2 insert was sequenced completely on both strands. pBNSF3-f2 nucleotide sequence is shown in SEQ ID NO:6. A comparison of this sequence with that of the Arabidopsis thaliana delta-15 desaturase clone (SEQ ID NO:1) confirmed that pBNSF3-f2 is a Brassica napus cDNA that encodes a seed microsomal delta-15 desaturase.

An additional low stringency screen of the Brassica napus seed cDNA library using the cDNA insert in pCM2 as a probe identified eight strongly-hybridizing phages. These phages were plaque purified and used to excise the phagemids as described above. One of these, designated pBNSFd-8, contained a 0.3kb insert. pBNSFd-8 was sequenced completely on one strand, this sequence had significant divergence from the sequence of pBNSF3-f2. The cDNA insert in pBNSFd-8 was used as a hybridization probe in a high stringency screen of the Brassica napus seed cDNA library. The filter hybridizations were carried out in 50 mM Tris pH 7.6, 6X SSC, 5X Denhardt's, 0.5% SDS, 100 ug denatured calf thymus DNA overnight at 50°C and post hybridization washes were in 6X SSC, 0.5% SDS at room temperature for 15 min, then with 2X SSC, 0.5% SDS at 45°C for 30 min, and then twice with 0.2X SSC, 0.5% SDS at 60°C for 30 min. The high stringency screen resulted in three strongly hybridizing phages that were purified and excised as above. One of the

excised plasmids pBNSFd-3 contained a 1.4kb insert that was sequenced completely on both strands. SEQ ID NO:8 shows the nucleotide sequence of pBNSFd-3. A comparison of this sequence with that of the Arabidopsis thaliana delta-15 desaturase clone (SEQ ID NO:4) confirmed that pBNSFd-3 is a Brassica napus cDNA that encodes a seed plastid delta-15 desaturase.

Cloning of a Soybean Seed cDNA Encoding a Microsomal Delta-15 Glycerolipid Desaturase

A cDNA library was made as follows: Soybean embryos (ca. 50 mg fresh weight each) were removed from the pods and frozen in liquid nitrogen. The frozen embryos were ground to a fine powder in the presence of liquid nitrogen and then extracted by Polytron homogenization and fractionated to enrich for total RNA by the method of Chirgwin et al. (Biochemistry (1979) 18:5294-5299). The nucleic acid fraction was enriched for poly A^+RNA by passing total RNA through an oligo-dT cellulose column and eluting the poly A^+RNA with salt as described by Goodman et al. (Meth. Enzymol. (1979) 68:75-90). cDNA was synthesized from the purified poly A^+RNA using cDNA Synthesis System (Bethesda Research Laboratory) and the manufacturer's instructions. The resultant double-stranded DNA was methylated by Eco RI DNA methylase (Promega) prior to filling-in its ends with T4 DNA polymerase (Bethesda Research Laboratory) and blunt-end ligation to phosphorylated Eco RI linkers using T4 DNA ligase (Pharmacia). The double-stranded DNA was digested with Eco RI enzyme, separated from excess linkers by passage through a gel filtration column (Sepharose CL-4B), and ligated to lambda ZAP vector (Stratagene) according to manufacturer's instructions. Ligated DNA was packaged into phage using the Gigapack packaging extract (Stratagene) according to manufacturer's instructions. The resultant cDNA library

was amplified as per Stratagene's instructions and stored at -80°C.

Following the instructions in the Lambda ZAP Cloning Kit Manual (Stratagene), the cDNA phage library was used to infect E. coli BB4 cells and approximately 80,000 plaque forming units were plated onto 150 mm diameter petri plates. Duplicate lifts of the plates were made onto nitrocellulose filters (Schleicher & Schuell). The filters were prehybridized in 25 mL of hybridization buffer consisting of 50mM Tris-HCl, pH 7.5, 1 M NaCl, 1% SDS, 5% dextran sulfate and 0.1 mg/mL denatured salmon sperm DNA (Sigma Chemical Co.) at 50°C for 2 h. Radiolabeled probe prepared from pCF3 as described above was added, and allowed to hybridize for 18 h at 50°C. The probes were washed twice at room temperature with 2X SSPE, 1% SDS for five minutes followed by washing for 5 min at 50°C in 0.2X SSPE, 1% SDS. Autoradiography of the filters indicated that there was one strongly hybridizing plaque, and approximately five weakly hybridizing plaques. The more strongly hybridizing plaque was subjected to a second round of screening as before, excepting that the final wash was for 5 min at 60°C in 0.2X SSPE, 1% SDS. Numerous, strongly hybridizing plaques were observed, and one, well-isolated from other phage, was picked for further analysis.

Following the Lambda ZAP Cloning Kit Instruction Manual (Stratagene), sequences of the pBluescript vector, including the cDNA inserts, from the purified phage was excised in the présence of a helper phage and the resultant phagemid was used to infect E. coli XL-1 Blue cells. DNA from the plasmid, designated pXF1, was made by the alkaline lysis miniprep procedure described in Sambrook et al. (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989) Cold Spring Harbor Laboratory

Press). The alkali-denatured double-stranded DNA from pXF1 was completely sequenced on both strands. The insert of pXF1 contained a stretch of 1783 nucleotides which contained an unknown open-reading frame and also contained a poly-A stretch of 16 nucleotides 3' to the open reading frame, from nucleotides 1767 to 1783, followed by an Eco RI restriction site. The 2184 bases that followed this Eco RI site contained a 1145 bp open reading frame which encoded a polypeptide of about 68% identity to, and colinear with, the Arabidopsis delta-15 desaturase polypeptide listed in SEQ ID No:2. The putative start methionine of the 1145 bp open-reading frame corresponded to the start methionine of the Arabidopsis microsomal delta-15 peptide and there were no amino acids corresponding to a plastid transit peptide 5' to this methionine. When the insert in pXF1 was digested with Eco RI four fragments were observed, fragments of approximately 370 bp and 1400 bp fragments, derived from the first 1783 bp of the insert in pXF1, and fragments of approximately 600 bp and 1600 bp derived from the the other 2184 nucleotides of the insert in pXF1. Only the 600 bp and 1600 bp fragments hybridized with probe derived from pCF3 on Southern blots. It was deduced that pXF1 contained two different cDNA inserts separated by an Eco RI site and the second of these inserts was a 2184 bp cDNA encoding a soybean microsomal delta-15 desaturase. The complete nucleotide sequence of the 2184 bp soybean microsomal delta-15 cDNA contained in plasmid pXF1 is listed in SEQ ID No:10.

Cloning of a Soybean Seed cDNA Encoding a Plastid Delta-15 Glycerolipid Desaturase Using Soybean Microsomal Delta-15 Desaturase cDNA as an Hybridization Probe

A 1.0 kb fragment of the coding region of the soybean microsomal delta-15 desaturase cDNA contained in

plasmid pXF1 was excised by digestion with the restricti n enzyme Hha I. This 1.0 Kb fragment was purified by agarose gel electrophoresis and radiolabeled with 32P as previously described. The radiolabeled fragment was used to screen 100,000 plaque-forming units of the the soybean cDNA library as described above. Autoradiography of the filters indicated that there were eight hybridizing plaques and these were subjected to a second round of screening. Sequences of the pBluescript vector from all eight of the purified phages, including the cDNA inserts, were excised in the presence of a helper phage and the resultant phagemids were used to infect E. coli XL-1 Blue cells. DNA from the plasmids was made by the alkaline lysis miniprep procedure described in Sambrook et al. (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989) Cold Spring Harbor Laboratory Press). Restriction analysis showed they contained inserts ranging from 1.0 kb to 3.0 kb in size. One of these inserts, designated pSFD-118bwp, contained an insert of about 1700 bp. The alkali-denatured double-stranded DNA from pSFD-118bwp was completely sequenced on both strands. The insert of pSFD-118bwp contained a stretch of 1675 nucleotides which contained an open-reading frame encoding a polypeptide of about 80% identity with, and colinear with, the Arabidopsis plastid delta-15 desaturase polypeptide listed in SEQ ID No:5. The open-reading frame also encoded amino acids corresponding to a plastid transit peptide at the 5' end of the open-reading frame. The transit peptide was colinear with, and shared some homology to, the transit peptide described for the Arabidopsis plastid delta-15 glycerolipid desaturase. Based on the homology to Arabidopsis plastid delta-15 glycerolipid desaturase and because of the presence of a plastid transit peptide, the cDNA contained in plasmid pSFD-118bwp was deduced to

be a soybean plastid delta-15 glycerolipid desaturase. The complete nucleotide sequ nce of the 1675 bp soybean plastid delta-15 glycerolipid desaturase cDNA is listed in SEQ ID NO:12.

EXAMPLE 6

CLONING OF cDNA SEQUENCES ENCODING FATTY ACID DESATURASES BY POLYMERASE CHAIN REACTION

Analysis of the deduced protein sequences of the different higher plant glycerolipid desaturases described in this invention reveals to those skilled in the art regions of the amino acid sequences that have been conserved among higher plants and between higher plants and cyanobacterial des A. These short stretches of amino acids can be used to design oligomers as primers for polymerase chain reactions. Two amino acid sequences that are highly conserved between the des A and plant delta-15 desaturases polypeptides are amino acid sequences 97-108 and 299-311 (SEQ ID NO:2). Polymerase chain reactions (PCRs) were performed using GeneAmp® RNA PCR Kit (Perkin Elmer Cetus) following manufacturer's protocols. In one PCR experiment, SEQ ID NOS:22 and 23 were used as sense primers and either SEQ ID NOS:24 and 25 or SEQ ID NOS:26 and 27 as antisense primers on poly A+ RNA purified from both *Arabidopsis* leaf and canola developing seeds. For this, ca. 100 ng of polyA+ RNA was isolated as described previously and reverse-transcribed using the kit using random hexamers. Then the cDNA was used in PCR using 64 pmoles each of SEQ ID NOS:22 and 23 as sense primers and either a mixture of 64 pmoles of SEQ ID NO:24 and 78 pmoles of SEQ ID NO:25 or a mixture 35 pmoles of SEQ ID NO:26 and 50 pmoles of SEQ ID NO:27 by the following program: a) 1 cycle of 2 min at 95°C and 15 C at 50°C, b) 30 cycles of 3 min at 65°C (extension), 1 min 20 sec at 95°C (denaturation), 2 min at 50°C (annealing), and c) 1

cycle of 7 min at 65°C. PCR products were analyzed by gel electrophoresis. All PCRs resulted in PCR products of the correct size (ca. 630 bp). The PCR products from Arabidopsis and canola were purified and used as radiolabeled hybridization probes to screen the Lambda Yes Arabidopsis cDNA library at low stringency, as described above. This led to the isolation of a pure phage, which was excised to give plasmid pYacp7. The cDNA insert in pYacp7 was partially sequenced. Its sequence showed that it encoded an incomplete desaturase polypeptide that was identical to another cDNA (in plasmid pFadx-2) isolated by low-stringency hybridization as described previously. The composite sequence derived from the partial sequences from the cDNA inserts in pFadx-2 and pYacp7 is shown in SEQ ID NO:16 and the polypeptide encoded by it in SEQ ID NO:17. As discussed previously, SEQ ID NO:17 is a putative plastid delta-15 desaturase. A full-length version of pYacp7 can be readily isolated using it has a hybridization probe.

Two additional conserved regions correspond to aminoacid residues 130 to 137 and 249 and 256 of SEQ ID NO:7 (Brassica napus glycerolipid desaturase delta-15). Degenerate oligomers were designed to these regions with additional nucleotides containing a restriction site for Bam H1 were added to the 5' ends of each oligonucleotide to facilitate subcloning of the PCR products. The nucleotide sequences of these oligonucleotides named F2-3 and F2-3c are shown in SEQ ID NO:18 and SEQ ID NO:19 respectively.

Mixtures of degenerate oligonucleotides F2-3 and F2-3c were used to amplify, isolate and clone glycerolipid desaturase sequences represented in corn seed mRNA population, essentially as described in the GeneAmp RNA PCR Kit purchased from Perkin Elmer Cetus and in Innis,

et al., Eds, (1990) PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego.

Corn seed RNA was obtained from developing corn seeds 15-20 days after pollination by the method of Chirgwin et al., (1979) Biochemistry 18:5294. Corn seed polyadenylated mRNA was isolated by affinity chromatography on oligo-dT cellulose (Aviv et al., Proc. Natl. Acad. Sci. USA (1972) 69:1408-1411). 20-50ng of A+mRNA were used in reverse transcription reactions with oligo-dT and random hexamers primers using the reaction buffer and conditions recomended by Perkin Elmer Cetus. The resulting cDNA was then used as template for the amplification of corn seed glycerolipid sequences using the set of degenerate primers in SEQ ID NO: 18 and 19. Reaction conditions were as described by Perkin Elmer Cetus, the amplification protocol consisted of a sequence of 95°C/1 min, 55°C/1 min, 72°C/2 min for 30-50 cycles. The resulting polymerase reaction products were phenol-chloroform extracted, digested with Bam HI and separated from unincorporated primers by gel filtration chromatopgraphy on Linker 6 spin columns (Pharmacia Inc.). The resulting PCR products were cloned into pBluescript SK at the Bam H1 site, and transformed into E. coli DH5 competent cells. Restriction analysis of plasmid DNA from the transformed colonies obtained revealed a colony, PCR-20, that contained an insert of about 0.5 kB in size at the pBluescript SK BamH1 site. The PCR-20 insert was completely sequenced on both strands. The nucleotide sequence of PCR20 insert is shown in SEQ ID NO:14 and the translated amino acid sequence is shown in SEQ ID NO:15. This aminoacid sequence shows an overall identity of 61.9% to the aminoacid sequence of Brassica napus microsomal delta-15 deaturase shown in SEQ ID NO:7. This result identifies the PCR20 insert as a polymerase reaction product of a

corn seed delta-15 desaturase cDNA. PCR20 insert may be used as a probe to readily isolate full length corn seed delta-15 desaturase cDNAs or as such to antisense or cosuppress corn seed glycerolipid delta-15 desaturase gene expression in transgenic corn plants by cloning it in the appropriate corn gene expression vector.

EXAMPLE 7

USE OF THE ARABIDOPSIS THALIANA DELTA-15 DESATURASE GENOMIC CLONES AS A RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP) MARKERS TO MAP THE DELTA-15 DESATURASE LOCI IN ARABIDOPSIS

DNA flanking the T-DNA insertion site in mutant line 3707 was used to map the genetic locus encoding the delta-15 desaturase of *Arabidopsis thaliana* seeds. An approximately 12 kB genomic DNA fragment containing the *Arabidopsis* delta-15 desaturase coding sequence was removed from the lambda-4211 clone by digestion with restriction endonuclease Xho I, separated from the Lambda arms by agarose gel electrophoresis, and purified using standard procedures. The isolated DNA was labeled with ^{32}P using a random priming kit from Pharmacia under conditions recommended by the manufacturer. The radioactive DNA was used to probe a Southern blot containing genomic DNA from *Arabidopsis thaliana* (ecotype Wassileskija and marker line W100 ecotype Landesberg background) digested with one of several restriction endonucleases. Following hybridization and washes under standard conditions (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed. (1989) Cold Spring Harbor Laboratory Press), autoradiograms were obtained. Different patterns of hybridization (polymorphisms) were identified in digests using restriction endonucleases Bgl II, Cla I, Hind III, Nsi I, and Xba I. The same radiolabeled DNA fragment was used to map the polymorphism essentially as described by

Helentjaris et al., (Theor. Appl. Genet. (1986) 72:761-769). The radiolabeled DNA fragment was applied as described above to Southern blots of Xba I digested genomic DNA isolated from 117 recombinant inbred progeny (derived from single-seed descent lines to the F_6 generation) resulting from a cross between Arabidopsis thaliana marker line W100 and ecotype Wassileskija (Burr et al., Genetics (1988) 118:519-526). The bands on the autoradiograms were interpreted as resulting from inheritance of either paternal (ecotype Wassileskija) or maternal (marker line W100) DNA or both (a heterozygote). The resulting segregation data were subjected to genetic analysis using the computer program Mapmaker (Lander et al., Genomics (1987) 1:174-181). In conjunction with previously obtained segregation data for 63 anonymous RFLP markers and 9 morphological markers in Arabidopsis thaliana (Chang et al., Proc. Natl. Acad. Sci. USA (1988) 85:6856-6860; Nam et al., Plant Cell (1989) 1:699-705), a single genetic locus was positioned corresponding to the genomic DNA containing the delta-15 desaturase coding sequence. The location of the delta-15 desaturase gene was thus determined to be on chromosome 2 between the lambda AT283 and cosmid c6842 RFLP markers, near the py and erecta morphological markers.

The cDNA in plasmid pCM2 was also shown to hybridize polymorphically to genomic DNA from Arabidopsis thaliana (ecotype Wassileskija and marker line W100 ecotype Landesberg background) digested with Eco RI. It was used as a RFLP marker to map the genetic locus for the gene encoding this fatty acid desaturase in Arabidopsis as described above. A single genetic locus was positioned corresponding to this desaturase cDNA. Its location was thus determined to be on chromosome 3 between the lambda AT228 and cosmid c3838

RFLP markers, "north" of the glabrous locus (Chang et al., Proc. Natl. Acad. Sci. USA (1988) 85:6856-6860; Nam et al., Plant Cell (1989) 1:699-705).

EXAMPLE 8

USE OF SOYBEAN SEED MICROSOMAL DELTA-15 GLYCEROLIPID DESATURASE cDNA SEQUENCE IN PLASMID AS A RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP) MARKER

A 600 bp fragment of the cDNA insert from plasmid pXF1, which contains about 300 bp of the coding sequence and 300 bp of the 3' untranslated sequence, was excised by digestion with restriction enzyme Eco RI in standard conditions as described in Sambrook et al. (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989) Cold Spring Harbor Laboratory Press), purified by agarose gel electrophoresis and labeled with ^{32}P using a Random Priming Kit from Bethesda Research Laboratories under conditions recommended by the manufacturer. The resulting radioactive probe was used to probe a Southern blot (Sambrook et al., Molecular Cloning, A Laboratory Manual, 2nd ed. (1989) Cold Spring Harbor Laboratory Press) containing genomic DNA from soybean [Glycine max (cultivar Bonus) and Glycine soja (PI81762)], digested with one of several restriction enzymes. After hybridization and washes under standard conditions (Sambrook et al. Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press), autoradiograms were obtained and different patterns of hybridization (polymorphisms) were identified in digests performed with restriction enzymes Bam HI, Eco RV and Eco RI. The same probe was then used to map the polymorphic pXF1 locus on the soybean genome, essentially as described by Helentjaris et al. (Theor. Appl. Genet. (1986) 72:761-769). Plasmid pXF1/600 bp probe was applied, as described above, to Southern blots of EcoRI, PstI, EcoRV, BamHI, or Hin DIII digested

genomic DNAs isolated from 68 F2 progeny plants resulting from a *G. max* Bonus x *G. soja* PI81762 cross. The bands on the autoradiograms were interpreted as resulting from the inheritance of either paternal (Bonus) or maternal (PI81762) pattern, or both (a heterozygote). The resulting data were subjected to genetic analysis using the computer program Mapmaker (Lander et al., Genomics (1987) 1:174-181). In conjunction with previously obtained data for 436 anonymous RFLP markers in soybean (Tingey et al., J. Cell. Biochem., Supplement 14E (1990) p. 291, abstract R153], Applicants were able to position a single genetic locus corresponding to the pXF1/600 bp probe on the soybean genetic map. This confirms that the gene for microsomal delta-15 desaturase is located on chromosome 19 in the soybean genome. This information will be useful in soybean breeding targeted towards developing lines with altered polyunsaturate levels.

EXAMPLE 9

OVEREXPRESSION OF MICROSOMAL DELTA-15 FATTY ACID DESATURASE IN PLANTS

Detailed procedures for DNA manipulation, such as use of restriction endonucleases and other DNA modifying enzymes, agarose gel electrophoresis, isolation of DNA from agarose gels, transformation of *E. coli* cells with plasmid DNA, and isolation and sequencing of plasmid DNA are described in Sambrook et al. (1989) Molecular cloning, A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press and Ausubel et al. (1989) Current Protocols in Molecular Biology John Wiley & Sons. All restriction enzymes and modifying enzymes were obtained from Bethesda Research Laboratory, unless otherwise noted.

To test the biological effect of overexpression of the microsomal delta-15 desaturase SEQ ID NO:1, i.e., the cDNA encoding *Arabidopsis thaliana* microsomal delta-15 desaturase, was placed in the sense orientation behind either the CaMV 35S promotor, to provide constituitive expression, or behind the promotor for the gene encoding soybean a' subunit of the β-conglycinin (7S) seed storage protein, to provide embryo-specific expression. To create the chimeric gene constructs, specific expression cassettes were made to facilitate easy manipulation of the desired clones. The chimeric genes were then transformed into plant cells by *Agrobacterium tumefaciens*'s binary Ti plasmid vector system [Hoekema et al,. (1983) Nature 303:179-180; Bevan (1984) Nucl. Acids Res. 12:8711-8720].

Overexpression of Arabidopsis Delta-15 Fatty Acid Desaturase in Transgenic Carrot Hairy Roots

To confirm the identity of SEQ ID NO:1 (Arabidopsis microsomal delta-15 fatty acid desaturase) and to test the biological effect of its overexpression in a heterologous plant species, the constitutive chimeric gene 35S:SEQ ID NO:1 was introduced into carrot tissue by Agrobacterium. The cassette for constitutive gene expression in plasmid, pAW28, originated from pK35K which, in turn, is derived from pKNK. Plasmid pKNK is a pBR322-based vector containing a chimeric gene for plant kanamycin resistance: nopaline synthase (NOS) promoter/neomycin phosphotransferase (NPT) II coding region/3' NOS chimeric gene. Plasmid pKNK has been deposited on 7 January 1987 with the American Type Culture Collection of Rockville,Maryland, USA under the provisions of the Budapest Treaty and bears the deposit accession number 67284. A map of this plasmid is shown in Lin, et al., Plant Physiol. (1987) 84:856-861. The NOS promoter region is a 296 bp Sau 3A-Pst I fragment

corresponding to nucleotides -263 to +33, with respect to th transcription start sit , of the NOS gene described by Depicker et al. (1982) J. Appl. Genet. 1:561-574. The Pst I site at the 3' end was created at the translation initiation codon of the NOS gene. The NptII coding region is a 998 bp Hind III-Bam HI fragment obtained from transposon Tn5 (Beck et al., Gene (1982) 19:327-336) by the creation of Hind III and Bam HI sites at nucleotides 1540 and 2518, respectively. The 3' NOS is a 702 bp Bam HI-Cla I fragment from nucleotides 848 to 1550 of the 3' end of the NOS gene (Depicker et al., J. Appl. Genet. (1982) 1:561-574) including its' polyadenylation region. pKNK was converted to pK35K by replacing its Eco RI-Hind III fragment containing the NOS promoter with a Eco RI-Hind III fragment containing the CaMV 35S promoter. The Eco RI-Hind III 35S promoter fragment is the same as that contained in pUC35K that has been deposited on 7 January 1987 with the American Type Culture Collection under the provisions of the Budapest Treaty and bears the deposit accession number 67285. The 35S promoter fragment was prepared as follows, and as described in Odell et al., Nature (1985) 313:810-813, except that the 3' end of the fragment includes CaMV sequences to +21 with respect to the transcription start site. A 1.15 KB Bgl II segment of the CaMV genome containing the region between -941 and +208 relative to the 35S transcription start site was cloned in the Bam HI site of the plasmid pUC13. This plasmid was linearized at the Sal I site in the polylinker located 3' to the CaMV fragment and the 3' end of the fragment was shortened by digestion with nuclease Bal31. Following the addition of Hind III linkers, the plasmid DNA was recircularized. From nucleotide sequence analysis of the isolated clones, a 3' deletion fragment was selected with the Hind III

linker positioned at +21. The 35S promoter fragment was isolated as an Eco RI-Hind III fragment, the Eco RI site coming from the polylinker of pUC13.

The NPTII coding region in plasmid pK35K was removed from plasmid pK35K by digestion with Hind III and Bam HI restriction enzymes. Following digestion, the ends of the DNA molecules were filled-in using Klenow enzyme. Not I linkers (New England Biolabs) were then ligated on the ends and the plasmid was recircularized to yield plasmid pK35Nt. A 1.7 kB fragment containing the 35S promotor region - Not I site - 3' untranslated region from nopaline synthase was liberated from pK35Nt using restriction endonucleases Eco RI and Cla I. Following restriction digestion the ends of the DNA molecules were filled-in using Klenow enzyme after which Xho I linkers (New England Biolabs) were added. The 1.7 kB fragment, now containing Xho I sites at either end, was gel isolated and cloned into the plasmid vector pURA3 (Clonetech) at its unique Xho I site. The vector pURA3 was choosen due to the absence of a Not I restriction site, the presence of a single Xho I restriction site and because the relatively large size of the vector (pURA3) would make the isolation of the gene expression cassettes relatively easy from the final construct.

The 1.4 kB Not I fragment in plasmid pCF3 containing Arabidopsis microsomal delta-15 desaturase (SEQ ID NO:1) was isolated and ligated to pAW28 (the constituitive expression cassette) previously linearized with Not I restriction enzyme and treated with calf intestinal alkaline phosphatase (Boehringer Mannheim) to result in plasmids pAW29 and pAW30 that had SEQ ID NO:1 cloned in a sense orientation and antisense orientation, respectively, with respect to the promoter. The orientation of the cDNA relative to the promotors was

established by digestion with appropriate restriction endonucleases or by sequencing across the promotor-cDNA junctions.

The chimeric genes 35S promotor/sense SEQ ID NO:1/3'NOS and 35S promotor/antisense SEQ ID NO:1/3'NOS were isolated as a 3 kB Xho I fragment from plasmids pAW29 and pAW30, respectively, and cloned into the binary vector pZS194b at its unique Sal I site to result in plasmids pAW31 and pAW32, respectively. The orientation of the plant selectable marker gene in pAW31 and pAW32 is the same as that of the 35S promoter as acertained by digestion with appropriate restriction endonucleases. Binary vector pZS194b contains the pBR322 origin of replication, the replication and stability regions of the _Pseudomonas aeruginosa_ plasmid pVS1 [Itoh, et al., (1984) Plasmid 11:206-220] required for replication and maintenance of the plasmid in _Agrobacterium_, the bacterial NPT II gene (kanamycin resistance) from Tn5 [Berg et al., (1975) Proc. Nat'l. Acad. Sci. U.S.A. 72:3628-3632] as a selectable marker for transformed bacteria, left and right borders of the T-DNA of the Ti plasmid [Bevan et al., (1984) Nucl. Acids Res. 12:8711-8720], and, between the left and right T-DNA borders are the chimeric NOS:NPT II gene for plant kanamycin resistance, described above, as a selectable marker for transformed plant cells and the E. coli lacZ a-complementing segment [Vieria and Messing (1982) Gene 19:259-267] with unique restriction endonuclease sites for Kpn I and Sal I.

The binary vectors pAW31 and pAW32 were transformed by the freeze/thaw method [Holsters et al. (1978) Mol. Gen. Genet. 163:181-187] into _Agrobacterium tumefaciens_ strain R1000, carrying the Ri plasmid pRiA4b from _Agrobacterium rhizogenes_ [Moore et al., (1979) Plasmid

2:617-626] to result in transformants R1000/pAW31 and R1000/pAW32, respectively.

Carrot (Daucus carota L.) cells were transformed by co-cultivation of carrot root disks with strain R1000, R1000/pAW31, or R1000/pAW32 by the method of Petit et al., (1986) Mol. Gen. Genet. 202:388-393]. To prepare explants for inoculation, carrots purchased from the local supermarket were first scrubbed gently with water and dish detergent, then rinsed thoroughly with tap and distilled water. They were surface sterilized in a stirred solution of 50% Clorox and distilled water for 30 min and rinsed thoroughly with sterile distilled water. The carrots were peeled using an autoclaved vegetable peeler and then sliced with a scalpel blade into disks of approximately 5-10 mm thickness. The disks were placed in petri dishes, onto a medium consisting of distilled deionized water solidified with 0.7% agar, in an inverted orientation so that the cut surface nearest to the root apex of the carrot was exposed for inoculation.

Cultures of Agrobacterium strains R1000, R1000/pAW31, and R1000/pAW32 were initiated from freshly grown plates in LB broth plus the appropriate antibiotic selective agents (50 mg/L chloramphenicol for the R1000 or 50 mg/L each of chloramphenicol and kanamycin for R1000/pAW31 and R1000/pAW32) and grown at 28°C to an optical density of around 1.0 at 600 nm. Bacterial cells were pelleted by centrifugation, rinsed and resuspended in LB broth without antibiotics. Freshly cut carrot disks were inoculated by applying 100 µL of the bacterial suspension to the cut surface of each disk. As a control, some disks were inoculated with sterile LB broth only, to indicate the extent of root formation in the absence of Agrobacterium.

Inoculated root disks were incubated at 25°C in the dark in petri dishes sealed with Parafilm. After two weeks of co-cultivation of carrot disks with Agrobacterium, the carrot disks were transferred to fresh agar-solidified water medium containing 500 mg/L carbenicillin for the counterselection of Agrobacterium. At this time, hairy root formation was noted on some root disks. Transfer of the explants to fresh counterselection medium was done at four weeks. Excision of individual roots from the explants was begun at six weeks. Ten days later, additional roots were taken from the explants as needed.

Approximately 5-10 mm long hairy roots were excised and individually subcultured on MS minimal organics medium with 30 g/L sucrose (Gibco, Grand Island, N. Y., Cat. No. 510-1118EA) and 500 mg/L carbenicillin. Approximately equal numbers of roots were subcultured in liquid medium and in a medium solidified with 0.6% agarose. Cultures on solid medium were grown in 60 x 100 mm petri dishes, liquid cultures were in 6-well culture dishes. When excising roots, an effort was made to select single roots from distinct callus-like outgrowths on the wounded surface. These sites of excision were marked on the lid of the petri dish to minimize repeat sampling of tissue originating from the same transformation event.

Two to three weeks after excision from the explants, individual hairy root cultures that were not visibly contaminated with Agrobacterium were transferred to fresh MS medium supplemented with 500 mg/L carbenicillin. The root mass of each culture was cut into segments including one or more branch roots, and these segments were transferred as a group to a plate or well of fresh medium. Approximately 20 mg fresh weight of tissue of root cultures which grew to adequate size

within the next two to three weeks were sampled for fatty acid composition by gas chromatography of the fatty acyl methyl esters essentially as described by Browse et al., (Anal. Biochem. (1986) 152:141-145) except that 2.5% H_2SO_4 in methanol was used as the methylation reagent and samples were heated for 1.5 h at 80°C to effect the methanolysis of the seed triglycerides. The results are shown in Table 6. A second sample of tissue consisting of an actively growing root tip of approximately 1 cm was excised and placed on MS medium supplemented with 500 mg/L carbenicillin and 25-50 mg/L kanamycin to test for kanamycin resistance select for hairy roots co-transformed with the binary vector [Simpson et al. (1986) Plant Mol. Biol. 6:403-415].

TABLE 6

Percent 18:3 and 18:2/18:3 Ratio in Roots of Transgenic Carrots

Root Sample	Transformation Vector Used	%18:3	%18:2/18:3
1	R1000/pAW31	62	0.09
2	R1000/pAW31	8	7.30
3	R1000/pAW31	10	5.69
4	R1000/pAW31	62	0.06
5	R1000/pAW31	10	5.07
6	R1000/pAW31	4	14.2
7	R1000/pAW31	61	0.18
8	R1000/pAW31	4	15.1
9	R1000/pAW31	61	0.07
10	R1000/pAW31	63	0.09
11	R1000/pAW31	15	3.04
12	R1000/pAW31	64	0.14
13	R1000/pAW31	5	9.94
14	R1000/pAW31	9	6.72
15	R1000/pAW31	8	7.08
16	R1000/pAW31	8	6.31
17	R1000/pAW31	23	1.86
18	R1000/pAW31	8	7.33
19	R1000/pAW31	10	5.99
20	R1000/pAW31	7	8.83
21	R1000/pAW32	9	6.80

Root Sample	Transformation Vector Used	%18:3	%18:2/18:3
22	R1000/pAW32	4	11.8
23	R1000/pAW32	3	18.8
24	R1000/pAW32	10	6.21
25	R1000/pAW32	7	8.57
26	R1000/pAW32	3	16.4
27	R1000/pAW32	6	8.29
28	R1000/pAW32	5	9.19
29	R1000/pAW32	5	8.47
30	R1000/pAW32	8	7.17
31	R1000/pAW32	4	11.9
32	R1000/pAW32	8	7.20
33	R1000/pAW32	5	10.4
34	R1000/pAW32	8	7.29
35	R1000/pAW32	3	17.2
36	R1000/pAW32	8	7.27
37	R1000/pAW32	9	6.01
38	R1000/pAW32	9	6.62
40	R1000/pAW32	9	6.02
41	R1000	8	7.23
42	R1000	8	7.83
43	R1000	10	6.20
44	R1000	9	5.97
45	R1000	9	6.73
46	R1000	9	6.27
47	R1000	8	7.27
48	R1000	7	8.30
49	R1000	9	7.11

The ability of R1000 transformed "hairy" roots to grow in the absence of exogenous phytohormones can be attributed to the Ri plasmid, pRiA4b. When R1000/pAW31 or R1000/pAW32 strains are used to transform, only a fraction (about half) of the "hairy" roots will also be

transformed with the experimental binary vector, pAW31 or pAW32. Thus, as expected, not all hairy roots resulting from transformation with R1000/pAW31 show the high 18:3 phenotype. The absense of any significant fatty acid phenotype in "hairy roots" transformed with R1000/pAW31 is expected, since carrot and Arabidopsis delta-15 desaturase sequences are not expected to be sufficiently related. These results show that overexpression of Arabidopsis microsomal delta-15 desaturase can result in over 10-fold increase in 18:3 at the expense of 18:2 in heterologous plant tissue.

Overexpression of Arabidopsis Delta-15 Fatty Acid Desaturase in Seeds and Complementation of the Mutation in Delta-15 Desaturation in Mutant 3707

To complement the delta-15 desaturation mutation in the T-DNA mutant 3707 and to test the biological effect of overexpression of SEQ ID NO:1 (Arabidopsis microsomal delta-15 fatty acid desaturase) in seed, the embryo-specific promoter:SEQ ID NO:1 chimeric gene was transformed into the mutant plant. This embryo-specific expression cassette in pAW42 was produced, in part, using a modified version of vector pCW109. Vector pCW109 itself was made by inserting into the Hind III site of the cloning vector pUC18 (Bethesda Research Laboratory) a 555 bp 5' non-coding region (containing the promoter region) of the β-conclycinin gene followed by the multiple cloning sequence containing the restriction endonuclease sites for Nco I, Sma I, Kpn I and Xba I, then 1174 bp of the common bean phaseolin 3' untranslated region into the Hind III site [Slightom et al., Proc. Nat'l Acad. Sci. U.S.A. (1983) 80:1897-1901]. The β-conclycinin promoter region used is an allele of the published β-conglycinin gene (Doyle et al., J. Biol. Chem. (1986) 261:9228-9238) due to differences at 27

nucl otide positions. Further sequence description may be found in Slightom (W091/13993).

The modifications to vector pCW109 were as follows: The potential translation start site was destroyed by digestion with Nco I and Xba I restriction enzymes followed by treatment with mung bean nuclease (New England Biolabs) to create linear, blunt ended DNA molecules. After ligation of Not I linkers (New England Biolabs) and digestion with Not I restriction enzyme (New England Biolabs) the plasmid was recircularized. Confirmation of the desired change was obtained by dideoxy sequencing. The resulting plasmid was designated pAW35. The 1.8 kB Hind III fragment from pAW35 containing the modified β-conclycinin promotor/3' phaseolin region was subcloned into the Hind III site in plasmid vector pBluescript SK$^+$ (Stratagene) creating plasmid pAW36. Plasmid pAW36 was linerized at its unique Eco RI site and ligated to Eco RI/Xho I adaptors (Stratagene). Following digestion with Xho I, the 1.7 kB Xho I fragment containing the β-conclycinin promotor/Not I site/3'-phaseolin untranslated region was cloned into the Xho I site in pURA3 vector (Clonetech). The resultant plasmid, pAW42, contains the seed specific expression cassette bordered by Xho I sites to facilitate cloning into modified T-DNA binary vectors and a unique Not I site to facilitate cloning of target cDNA sequences. Vector pURA3 was choosen due to the absence of a Not I restriction site, the presence of a single Xho I restriction site, and the relatively large size of the vector (pURA3) would make the isolation of the gene expression cassettes relatively easy from the final construct.

The 1.4 kB Not I fragment in plasmid pCF3 containing Arabidopsis microsomal delta-15 desaturase (SEQ ID NO:1) was isolated and ligated to plasmid pAW42

(the seed-specific expression cassette) that had pr viously been linearized with Not I restriction enzyme and treated with calf intestinal alkaline phosphatase (Boehringer Mannheim) to result in plasmids pAW45 that had SEQ ID NO:1 cloned in a sense orientation with respect to the promoter. The orientation of the cDNA relative to the promotors was established by digestion with appropriate restriction endonucleases or by sequencing across the promotor-cDNA junctions.

The chimeric β-conclycinin promotor/sense SEQ ID NO:1/phaseolin 3' was isolated as a 3.2 kB Xho I fragment from plasmid pAW45 and subcloned into the binary vector pAW25 at its unique Sal I site. In the resulting vector, pAW50, the orientation of the plant selectable marker is the same as that of the β-conclycinin promoter as acertained by digestion with appropriate restriction endonucleases. Plasmid pAW25, is derived from plasmids pZS94K and pML2. Plasmid pZS94K contains the pBR322 origin of replication, the replication and stability regions of the Pseudomonas aeruginosa plasmid pVS1 [Itoh, et al., (1984) Plasmid 11:206-220] required for replication and maintenance of the plasmid in Agrobacterium, the bacterial NPT II gene (kanamycin resistance) from Tn5 [Berg et al., (1975) Proc. Nat'l. Acad. Sci. U.S.A. 72:3628-3632] as a selectable marker for transformed bacteria, a T-DNA left border fragment of the octopine Ti plasmid pTiA6 and right border fragment derived from TiAch5 describe by van den Elzen et al. (Plant Mol. Biol. (1985) 5:149-154). Between these borders are the E. coli lacZ a-complementing segment [Vieria and Messing (1982) Gene 19:259-267] with restriction endonuclease sites Sal I and Asp 718 derived from pUC18. A 4.5 kB Asp 718-Sal I DNA fragment containing the chimeric herbicide sulfonylurea (SU)-resistant acetolactate (ALS) gene was

obtained from plasmid pML2 and cloned into the Asp 718-Sal I sites of plasmid pZS94K. This chimeric ALS gene contained the CaMV 35S promoter/Cab22L Bgl II-Nco I fragment that is described by Harpster et al., [Mol. Gen. Genet. (1988) 212:182-190] and the Arabidopsis ALS coding and 3' non-coding sequences [Mazur et al., (1987) Plant Physiol. 85:1110-1117] that was mutated so that it encodes a SU-resistant form of ALS. The mutation, introduced by site-directed mutagenesis, are those present in the tobacco SU-resistant Hra gene described by Lee et al., (1988) EMBO J. 5:1241-1248. The resulting plasmid was designated pAW25.

The binary vector pAW25 containing the chimeric embryo-specific β-conglycinin promotor:sense SEQ ID NO:1 gene was transformed by the freeze/thaw method [Holsters et al., (1978) Mol. Gen. Genet. 163:181-187] into the avirulent Agrobacterium strain LBA4404/pAL4404 [Hoekema et al., (1983) Nature 303:179-180].

Arabidopsis root cultures were transformed by co-cultivation with Agrobacterium using standard aseptic techniques for the manipulation of sterile media and axenic plant/bacterial cultures were followed, including the use of a laminar flow hood for all transfers. Compositions of the culture media are listed in Table 8. Unless otherwise indicated, 25x100 mm petri plates were used for plant tissue cultures. Incubation of plant tissue cultures was at 23°C under constant illumination with mixed fluorescent and "Gro and Sho" plant lights (General Electric) unless otherwise noted. To initiate *in vitro* root cultures of the T-DNA homogyzous mutant line 3707 (*Arabidopsis thaliana* (L.) Heynh, geographic race Wassilewshija) seeds of the mutant line were sterilized for 10 min in a solution of 50% Chlorox with 0.1% SDS, rinsed 3 to 5 times with sterile dH_2O, dried thoroughly on sterile filter paper, and then 2-3 seeds

were sown in liquid B5 medium in 250 mL Belco flasks. The flasks were capped, placed on a rotary shaker at 70-80 rpm, and incubated for 3-4 weeks. Prior to inoculation with Agrobacterium, root tissues were cultured on callus induction medium (MSKig). Roots were harvested by removing the root mass from the Belco flask, placing it in a petri dish, and, using forceps, pulling small bundles of roots from the root mass and placing them on MSKig medium. Petri dishes were sealed with filter tape and incubated for four days.

Agrobacterium strain LBA4404 carrying the plasmids pAL4404 and pAW50 were grown in 5 mL of YEB broth containing 25 mg/L kanamycin and 100 mg/L rifampicin. The culture was grown for approximately 17-20 h in glass culture tubes in a New Brunswick platform shaker (225 rpm) maintained at 28°C. Pre-cultured roots were cut into 0.5 cm segments and placed in a 100 µm filter, made from a Tri-Pour beaker (VWR Scientific, San Francisco, CA USA) and wire mesh, which is set in a petri dish. Root segments were inoculated for several min in 30-50 mL of a 1:20 dilution of the overnight Agrobacterium culture with periodic gentle mixing. Inoculated roots were transferred to sterile filter paper to draw off most of the liquid. Small bundles of roots, consisting of several root segments, were placed on MSKig medium containing 100 µM acetosyringone (3',5'-Dimethoxy-4'-hydroxyaceto-phenone, Aldrich Chemical Co., Milwaukee, WI, USA). Petri plates were sealed with parafilm or filter tape and incubated for 2 to 3 days.

After infection, root segments were rinsed and transferred to shoot induction medium with antibiotics. Root bundles were placed in a 100 µm filter unit (described above) and rinsed with 30-50 mL liquid MSKig medium. The filter was vigorously shaken in the solution to help remove the Agrobacterium, transferred

to a clean petri dish, and rinsed again. Roots were blotted on sterile filter paper and bundles of roots were placed on MSg medium containing 500 mg/l vancomycin and either 10 or 20 ppb chlorsulfuron. Plates were sealed with filter tape and incubated for 12 to 14 days.

Green nodules and small shoot primordia were visible at about 2-3 weeks. The explants were either left intact or were broken into numerous pieces and placed on GM medium containing 200-300 mg/L vancomycin and either 10 or 20 ppb chlorsulfuron for further shoot development. Plates were either sealed with two pieces of tape or with filter tape. As they developed, individual shoots were isolated from the callus and were placed on MSRg medium containing 100 mg/L vancomycin and either 10 or 20 ppb chlorsulfuron. Dishes were sealed as described above and incubated for seven to 10 days. Shoots were then transferred to GM medium containing 100-200 mg/L vancomycin in 25x100 petri dishes or Magenta G7 vessels. Many primary transformants (T1) which were transferred to individual containers set seed (T2).

T2 seed was harvested from selected putative transformants and sown on GM medium containing 10ppb chlorsulfuron. Plates were sealed with filter tape, cold treated for 2 or more days at 4°C, and then incubated for 10 to 20 days at 23°C under constant illumination as described above. Seedlings were scored as resistant (green, true leaves develop) and sensitive (no true leaves develop).

Selected chlorsulfuron resistant T2 seedlings were transplanted to soil and were grown to maturity at 23°C daytime (16 h) 18°C nighttime (8 h) at 65-80% relative humidity.

T2 seeds from two plants were harvested at maturity and analysed individually for fatty acid composition by

gas chromatography of the fatty acyl methyl esters essentially as described by Browse et al., (Anal. Biochem. (1986) 152:141-145) except that 2.5% H_2SO_4 in methanol was used as the methylation reagent and samples were heated for 1.5 h at 80°C to effect the methanolysis of the seed triglycerides. The results are shown in Table 7.

TABLE 7

Percent Fatty Acid in Seeds of Transgenic Mutant 3707

Seed Sample	16:0	18:0	18:1	18:2	18:3
wildtype (6)	6	4	14	30	19
mutant 3707 (6)	6	4	14	44	3
1-1	10	4	22	9	55
1-2	11	6	22	14	48
1-3	12	7	16	6	57
1-4	10	4	30	52	4
1-5	10	4	18	17	48
1-6	10	5	15	15	53
2-1	11	5	19	60	4
2-2	10	5	19	9	56
2-3	9	4	27	8	52
2-4	10	5	17	10	56
2-5	10	5	19	9	56
2-6	10	5	17	17	48

The fatty acid composition of the wild-type and mutant line 3707 represents the average of 6 single seeds each. Seeds from plant 1 are designated 1-1 to 1-6 and those from plant 2 are designated 2-1 to 2-6. The 20:1 and 20:2 amounts are not shown. The data shows that the one out of six seeds in each plant show the mutant fatty acid phenotype, while the remaining seeds show more than 10-fold increase in 18:3 to ca.55%. While most of the increase occurs at the expense of 18:2, some of it also occurs at the expense of 18:1.

Such high levels are of linolenic acid in vegetable oils are observed in specialty oil crops, such as linseed. Thus, overexpression of this gene in other oilscrops, especially canola, which is a close relative of Arabidopsis, is also expected to result in such high levels of 18:3.

TABLE 8

Medium Composition

YEP MEDIUM	
Bacto Beef Extract	5.0 g
Bacto Yeast Extract	1.0 g
Peptone	5.0 g
Sucrose	5.0 g
$MgSO_4 \cdot 7H_2O$	0.5 g
Agar (optional)	15.0 g
pH	
VITAMIN SUPPLEMENT	
10 mg/L thiamine	
50 mg/L pyridoxine	
50 mg/L nicotinic acid	
MSKIg = Callus Induction Medium	
Basic Medium	
2% glucose	20 g/L
0.5 mg/L 2,4-D	2.3 µL
0.3 mg/L Kinetin	1.4 µM
5 mg/L IAA	28.5 µM
MSRg = Shoot Induction Medium	
Basic Medium	
2% glucose	20 g/L
12 mg/L IBA	58.8 µM
0.1 mg/L Kinetin	0.46·µM

BASIC MEDIUM	
1 Pkg. Murashige and Skoog Minimal Organics Medium without Sucrose (Gibco #510-3118 or Sigma #M6899)	
10 mL Vitamin Supplement	
0.05% MES	0.5 g/L
0.8% agar	8 g/L
pH	
GM = Germination Medium	
Basic Medium	
1% sucrose	10 g/L
MSg = Shoot Induction Medium	
Basic Medium	
2% glucose	20 g/L
0.15 mg/L IAA	0.86 µM
5.0 mg/L 2iP	24.6 µM

EXAMPLE 10

Construction of Vectors for Transformation of *Brassica napus* for Reduced Expression of Delta-15 Desaturases in Developing Seeds

Detailed procedures for manipulation of DNA fragments by restriction endonuclease digestion, size separation by agarose gel electrophoresis, isolation of DNA fragments from agarose gels, ligation of DNA fragments, modification of cut ends of DNA fragments and transformation of *E. coli* cells with circular DNA plasmids are all described in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed (1989) Cold Spring Harbor Laboratory Press) and Ausubel et al., Current Protocols in Molecular Biology (1989) John Wiley & Sons).

Sequences of the cDNA's encoding the *B. napus* cytoplasmic delta-15 desaturase and the *Brassica napus* plastid delta-15 desaturase were placed in the antisense orientation behind the promoter region from the a' subunit of the soybean storage protein β-conglycinin to provide embryo specific expression and high expression levels.

An embryo-specific expression cassette was constructed to serve as the basis for chimeric gene constructs for anti-sense expression of the nucleotide sequences of delta-15 desaturase cDNAs. The vector pCW109 was produced by the insertion of 555 base pairs of the β-conglycinin (a' subunit of the 7s seed storage protein) promoter from soybean (*Glycine max*), the β-conglycinin 5' untranslated region followed by a multiple cloning sequence containing the restriction endonuclease sites for Nco I, Sma I, Kpn I and Xba I, then 1174 base pairs of the common bean phaseolin 3' untranslated region into the Hind III site in the cloning vector pUC18 (BRL). The β-conglycinin promoter

sequence represents an allele of the published β-conglycinin gene (Doyle et al., (1986) J. Biol. Chem. 261:9228-9238) due to differences at 27 nucleotide positions. Further sequence description may be found in Slightom (WO91/13993). The sequence of the 3' untranslated region of phaseolin is described in (Slightom et al.,(1983) Proc. Natl. Acad. Sci. USA, 80:1897-1901).

To facilitate use in antisense constructions, the Nco I site and potential translation start site in the plasmid pCW109 was destroyed by digestion with Nco I, mung bean exonuclease digestion and re-ligation of the blunt site to give the modified plasmid pCW109A. pCW109A was opened between the β-conglycinin promoter sequence and the phaseolin 3' sequence by digestion with Sma I to allow insertion of blunt ended cDNA fragments encoding the delta-15 desaturase sequences by ligation. The blunt ended fragment of the cytoplasmic delta-15 desaturase was obtained from plasmid pBNSF3, which contains the nucleotides 208 to 1336 of the cDNA insert described in SEQ ID NO:6. pBNSF3 was modified to remove the Hind III site at bases 682 to 687 of SEQ ID 6 by digesting with Hind III, blunting with Klenow and re-ligating. The resulting plasmid [pBNSF3(-H)], was digested with Eco RI and Xho I to release the delta-15 cDNA fragment, all ends were Klenow blunted and the 1.2 kB coding region was purified by gel isolation. The 1.2 kB fragment was ligated into the Sma I cut pCW109A described above. The antisense orientation of the inserted cDNA relative to the β-conglycinin promoter was established by digestion with Aat I which cuts in the delta-15 desaturase coding region and in the vector 5' to the β-conglycinin promoter to release a 1.4 Kb fragment when the coding region is in the antisense

orientation. The antisense construction was given the name pCCFdR1.

The transcription unit [β-conglycinin promoter:antisense delta-15 desaturase:phaseolin 3'end] was released from pCCFdR1 by Hind III digestion, isolated,and ligated into pBluescript which had also been Hind III digested to give plasmid pCCFdR2. This construct has unique BamH I and Sal I sites which were digested. The 3 kB transcriptional unit was isolated and cloned into the Bam HI and Sal I sites in pZ199 described below to give the binary vector pZCC3FdR. The orientation given by this directional cloning is with transcription of both the selectable marker gene and the delta-15 antisense gene in the same direction and toward the right border tDNA sequence.

An antisense construction based on the plastid delta-15 desaturase was made with the 425 most 3' bases of SEQ ID NO:8 which is contained in the plasmid pBNSFD-8. pBNSFD-8 represents a cDNA of the plastid delta-15 desaturase in pBluescript. The cDNA insert was removed from pBNSFD-8 by digestion with Xho I and Sma I, the fragments were blunted, and the 425 base insert isolated by gel purification. The isolated fragment was cloned into the Sma I site of pCW109A and the antisense orientation of the chosen clone confirmed by digestion of the plasmid with Pst I. Pst I cuts in the plastid delta-15 sequence and in the pCW109A vector 5' to the β-conglycinin promoter to release a 1.2 kB fragment indicative of the antisense orientation. The plasmid containing this construction was called pCCdFdR1.

Digestion of pCCdFdR1 with Hind III removes a 2.3 kB fragment containing the transcriptional unit [β-conglycinin promter:plastid delta-15 antisense:3'-phaseolin sequence]. The fragment was gel isolated and cloned into Hind III digested pBluescript. The

orientation of the fragment was relative to the Bam HI site in the cloning region of pBluescript was determined by digestion with Pst I as described above. A clone oriented with the promoter toward the Sal I containing end was chosen and given the name pCCdFdR2.

pCCdFdR2 was digested with Bam HI and Sal I, the released fragment was gel isolated and ligated into pZ199 which had been digested with Bam HI and Sal I to give the binary vector pZCCdFdR.

Vectors for transformation of the antisense delta-15 desaturase constructions under control of the β-conglycinin promoter into plants using Agrobacterium tumefaciens were produced by constructing a binary Ti plasmid vector system (Bevan, (1984) Nucl. Acids Res. 12:8711-8720). The starting vector used for these systems (pZS199) is based on a vector which contains: (1) the chimeric gene nopaline synthase/neomycin phosphotransferase as a selectable marker for transformed plant cells (Bevan et al., (1984) Nature 304:184-186), (2) the left and right borders of the T-DNA of the Ti plasmid (Bevan et al., (1984) Nucl. Acids Res. 12:8711-8720), (3) the E. coli lacZ a-complementing segment (Vieria and Messing (1982) Gene 19:259-267) with unique restriction endonuclease sites for Eco RI, Kpn I, Bam HI, Hin DIII, and Sal I, (4) the bacterial replication origin from the Pseudomonas plasmid pVS1 (Itoh et al., (1984) Plasmid 11:206-220), and (5) the bacterial neomycin phosphotransferase gene from Tn5 (Berg et al., (1975) Proc. Natnl. Acad. Sci. U.S.A. 72:3628-3632) as a selectable marker for transformed A. tumefaciens. The nopaline synthase promoter in the plant selectable marker was replaced by the 35S promoter (Odell et al. (1985) Nature, 313:810-813) by a standard restriction endonuclease digestion and ligation strategy. The 35S promoter is

required for efficient *Brassica napus* transformation as described below.

EXAMPLE 11

AGROBACTERIUM MEDIATED TRANSFORMATION OF *BRASSICA NAPUS*

The binary vectors pZCC3FdR abd pZCCdFdR were transferred by a freeze/thaw method (Holsters et al., (1978) Mol Gen Genet 163:181-187) to the *Agrobacterium* strain LBA4404/pAL4404 (Hoekema et al., (1983), Nature 303:179-180).

Brassica napus cultivar "Westar" was transformed by co-cultivation of seedling pieces with disarmed *Agrobacterium tumefaciens* strain LBA4404 carrying the the appropriate binary vector.

B. napus seeds were sterilized by stirring in 10% Chlorox, 0.1% SDS for thirty min, and then rinsed thoroughly with sterile distilled water. The seeds were germinated on sterile medium containing 30 mM $CaCl_2$ and 1.5% agar, and grown for six days in the dark at 24°C.

Liquid cultures of *Agrobacterium* for plant transformation were grown overnight at 28°C in Minimal A medium containing 100 mg/L kanamycin. The bacterial cells are pelleted by centrifugation and resuspended at a concentration of 10^8 cells/mL in liquid Murashige and Skoog Minimal Organic medium containing 100 µM acetosyringone.

B. napus seedling hypocotyls were cut into 5 mm segments which were immediately placed into the bacterial suspension. After 30 min, the hypocotyl pieces were removed from the bacterial suspension and placed onto BC-12 callus medium containing 100 µM acetosyringone. The plant tissue and *Agrobacteria* were co-cultivated for three days at 24°C in dim light.

The co-cultivation was terminated by transferring the hypocotyl pieces to BC-12 callus medium containing

200 mg/L carbenicillin to kill the *Agrobacteria*, and 25 mg/L kanamycin to select for transformed plant cell growth. The seedling pieces were incubated on this medium for three weeks at 24°C under continuous light.

After three weeks, the segments wre transferred to BS-48 regeneration medium containing 200 mg/L carbenicillin and 25 mg/L kanamycin. Plant tissue was subcultured every two weeks onto fresh selective regeneration medium, under the same culture conditions described for the callus medium. Putatively transformed calli grow rapidly on regeneration medium; as calli reached a diameter of about 2 mm, they were removed from the hypocotyl pieces and placed on the same medium lacking kanamycin.

Shoots began to appear within several weeks after transfer to BS-48 regeneration medium. As soon as the shoots formed discernable stems, they were excised from the calli, transferred to MSV-1A elongation medium, and moved to a 16:8 h day/night photoperiod at 24°C.

Once shoots had elongated several internodes, they were cut above the agar surface and the cut ends were dipped in Rootone. Treated shoots were planted directly into wet Metro-Mix 350 soiless potting medium. The pots were covered with plastic bags which were removed when the plants were clearly growing -- after about 10 days.

Plants were grown under a 16:8 h day/night photoperiod, with a daytime temperature of 23°C and a nighttime temperature of 17°C. When the primary flowering stem began to elongate, it was covered with a mesh pollen-containment bag to prevent outcrossing. Self-pollination was facilitated by shaking the plants several times each day. Seeds derived from self-pollinations were harvested about three months after planting.

TABLE 9

Minimal A Bacterial Gr wth Medium

Dissolve in distilled water:

10.5 g potassium phosphate, dibasic

4.5 g potassium phosphate, monobasic

1.0 g ammonium sulfate

0.5 g sodium citrate, dihydrate

Make up to 979 mLs with distilled water

Autoclave

Add 20 mLs filter-sterilized 10% sucrose

Add 1 mL filter-sterilized 1 M $MgSO_4$

Brassica Callus Medium BC-12

Per liter:

Murashige and Skoog Minimal Organic Medium (MS salts, 100 mg/L i-inositol, 0.4 mg/L thiamine; GIBCO #510-3118)

30 sucrose

18 g mannitol

1.0 mg/L 2,4-D

3.0 mg/L kinetin

0.6% agarose

pH 5.8

Brassica Regeneration Medium BS-48

Murashige and Skoog Minimal Organic Medium Gamborg B5 Vitamins (SIGMA #1019)

10 g glucose

250 mg xylose

600 mg MES

0.4% agarose

pH 5.7

Filter-sterilize and add after autoclaving:

2.0 mg/L zeatin

0.1 mg/L IAA

Brassica Shoot Elongation Medium MSV-1A

Murashige and Skoog Minimal Organic Medium Gamborg B5 Vitamins

10 g sucrose

0.6% agarose

pH 5.8

EXAMPLE 12

ANALYSIS OF TRANSGENIC BRASSICA NAPUS PLANTS

Insertion of the intact antisense transcriptional unit was verified by Southern analysis using transgenic plant leaf tissue as the source of DNA as described in Example 5. Ten micrograms of leaf DNA was digested to completion with a mixture of Bam HI and Sal I

restriction endonucleases and then separatd by agarose gel electrophoresis. The separated DNA was transferred to Hybond H^+ membrane and hybridized with radiolabeled insert from pBNSF3-2. An estimate of the number of copies of the inserted transgene was made by calibrating each Southern blot with standard amounts of pBNSF3-2 corresponding to 1 and 5 copies per genome and comparing intensities of the autoradiographic signal from the standards, the endogenous delta-15 desaturase signals and the inserted gene signal. To date, 38 independent transformants have been analyzed for presence of the gene and 36 were found to be positive.

The relative content of the 5 most abundant fatty acids in canola seeds was determined either by direct trans-esterification of individual seeds in 0.5 mL of methanolic H_2SO_4 (2.5%) or by hexane extraction of bulk seed samples followed by trans-esterification of an aliquot in 0.8 mL of 1% sodium methoxide in methanol. Fatty acid methyl esters were extracted from the methanolic solutions into hexane after the addition of an equal volume of water.

The relative content of 18:3 fatty acid varies significantly during seed development. To a lesser extent, the ratio of 18:3 to 18:2 varies also. Thus meaningful data can be obtained only from seeds after maturation and drydown. Additionally, the ratio of 18:3 to total fatty acid content and to 18:0 varies significantly due to environmental factors, primarily temperature. In this circumstance, the most appropriate controls are the transformed plants which by Southern analysis do not contain the antisense delta-15 transgene. Analysis from the first 5 transformants to reach dry seed are given in Table 10 below. Seeds were harvested using a hand thesher, bulked and a 1.5 g (about 300 seeds) sample was taken. Seed from each

transformant was crushed with a mortar and pestel, extracted 4 times with 8 mL hexane at about 50°C. The combined extracts were reduced in volume to 5 mL and two 50 microliter aliquots were taken for esterification as described above. Separation of the fatty acid methyl esters was done by gas-liquid chromatography using an Omegawax 320 column (Supelco Inc., 0.32 mm ID X 30M) run isothermally at 220° and cycled to 260° between each injection.

TABLE 10

Transformant No.	% 18:3	%18:3/18:2	Antisense delta-15 Copy No.
pZCC3FdR-91	6.2	0.39	0
pZCC3FdR-81	5.9	0.33	1
pZCC3FdR-15	6.0	0.38	2
pZCC3FdR-11	5.6	0.34	1
pZCC3FdR-148	8.2	0.40	2

The differences between the 4 transformed lines and line 92 are very small, however to test the significance of the difference in the 18:3/18:2 ratio between line 81 and 91, 25 individual seeds from each line were trans-esterified and their fatty acid composition determined. The average ratio for line 81 was 0.345 with a coefficient of variation of 11.6% while the average for line 91 was 0.375 with a coefficient of variation of 8.0%. The sample means are significantly different at the 0.01% level using Student's t test.

EXAMPLE 13

CONSTRUCTION OF VECTORS FOR TRANSFORMATION OF GLYCINE MAX FOR REDUCED EXPRESSION OF DELTA-15 DESATURASES IN DEVELOPING SEEDS

The antisense G. max plastid delta-15 desaturase cDNA sequence under control of the β-conglycinin promoter was constructed using the vector pCW109A described in Example 10 above. For use in the soybean

transformation system described below, the transcriptional unit was placed in a vector along with an appropriate selectable marker expression system. The starting vector was pML45, which consists of the non-tissue specific and constitutive promoter designated 508D and described in Hershey (WO 9011361) driving expression of the neomycin phosphotransferase gene described in (Beck et al. (1982) Gene 19:327-336) followed by the 3' end of the nopaline synthase gene including nucleotides 848 to 1550 described by (Depicker et al. (1982) J. Appl. Genet. 1:561-574). This transcriptional unit was inserted into the commercial cloning vector pGEM9Z (BRL) and is flanked at the 5' end of the 508D promoter by the restriction sites Sal I, Xba I, Bam HI and Sma I in that order. An additional Sal I site is present at the 3' end of the NOS 3' sequence and the Xba I, Bam HI and Sal I sites are unique.

Removal of the unit[β-conglycinin promter:cloning region:phaseolin 3' end] from pCW109A by digestion with Hind III, blunting the ends and isolating the 1.8 kB fragment afforded the expression cassette pCST by ligating the above isolated fragment into the Sma I site of pML45. A clone with the β-conglycinin promoter in the same orientation as the 508D promoter were chosen by digestion with Xba I. The correct orientation releases a 700 bp fragment. This vector cassette was called pCST.

The 2.2 kB insert encoding the soybean, plastid delta-15 desaturase was subcloned from the plasmid pXF1 by digestion with HinP I to remove about 1 kB of unrelated cDNA. HinP I cuts within the cDNA insert very near the 5' end of the cDNA for the delta-15 desaturase and about 300 bp from the 3' end of that cDNA. The Cla I compatable ends were cloned into Cla I digested pBluescript and a clone with the 5' end of the cDNA

toward the Eco RV site in the pBluescript cloning region was selected based on the relaese of a 900 bp fragment by digestion with Pst I. The subcloned plasmid was called pS3Fd1.

The delta-15 encoding sequence was removed from pS3Fd1 by digestion with HinC II and Eco RV, the 2.2 kB fragment was gel isolated and cloned into the opened Sma I site in pCST1. A clone with the delta-15 sequence in the antisense orientation to the β-conglycinin promoter was selected by digestion with Xba I. The antisense construct releases a 400 bp piece and that clone was designated pCS3FdST1R.

EXAMPLE 14

TRANSFORMATION OF SOMATIC SOYBEAN EMBRYO CULTURES

Soybean embryogenic suspension cultures are maintained in 35 mL liquid media (SB55 or SBP6) on a rotary shaker, 150 rpm, at 28°C with mixed florescent and incandescent lights on a 16:8 h day/night schedule. Cultures were subcultured every four weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures were transformed with pCS3FdST1R by the method of particle gun bombardment (see Kline et al. (1987) Nature (London) 327:70). A Du Pont Biolistic PDS1000/HE instrument (helium retrofit) was used for these transformations.

To 50 mL of a 60 mg/mL 1 mm gold particle suspension was added (in order); 5 µL DNA(1 µg/µL), 20 µL spermidine (0.1M), and 50 µl $CaCl_2$ (2.5 M). The particle preparation was agitated for 3 min, spun in a microfuge for 10 sec and the supernatant removed. The DNA-coated particles were then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension was sonicated three times

for 1 sec each. Five µL of the DNA-coated gold particles w re then loaded on each macro carrier disk.

Approximately 300-400 mg of a four week old suspension culture was placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue were normally bombarded. Membrane rupture pressure was set at 1000 psi and the chamber was evacuated to a vacuum of 28 inches of mercury. The tissue was placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue was placed back into liquid and cultured as described above.

Eleven days post bombardment, the liquid media was exchanged with fresh SB55 containing 50 mg/mL hygromycin. The selective media was refreshed weekly. Seven weeks post bombardment, green, transformed tissue was observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue was removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Thus each new line was treated as independent transformation event. These suspensions can then be maintained as suspensions of embryos clustered in an immature developmental stage through subculture or regenerated into whole plants by maturation and germination of individual somatic embryos.

Transformed embryogenic clusters were removed from liquid culture and placed on a solid agar media (SB103) containing no hormones or antibiotics. Embryos were cultured for eight weeks at 26°C with mixed florescent and incandescent lights on a 16:8 h day/night schedule. During this period, individual embryos were removed from the clusters and analyzed at various stages of embryo

development After eight weeks th embryos become suitable for germination.

TABLE 11

Media:

SB55 and SBP6 Stock Solutions (g/L):

MS Sulfate 100X Stock

$MgSO_4$ 7H2O	37.0
$MnSO_4$ H2O	1.69
$ZnSO_4$ 7H2O	0.86
$CuSO_4$ 5H2O	0.0025

MS Halides 100X Stock

$CaCl_2$ $2H_2O$	44.0
KI	0.083
$CoCl_2$ $6H_2O$	0.00125
KH_2PO_4	17.0
H_3BO_3	0.62
Na_2MoO_4 $2H_2O$	0.025

MS FeEDTA 100X Stock

Na_2EDTA	3.724
$FeSO_4$ $7H_2O$	2.784

B5 Vitimin Stock

10 g m-inositol

100 mg nicotinic acid

100 mg pyridoxine HCl

1 g thiamine

SB55 (per Liter)

10 mL each MS stocks

1 mL B5 Vitaimin stock

0.8 g NH_4NO_3

3.033 g KNO_3

1 mL 2,4-D (10mg/mL stock)

60 g sucrose

0.667 g asparagine

pH 5.7

For SBP6- substitute 0.5 mL 2,4-D

SB103 (per Liter)

MS Salts

6% maltose

750 mg $MgCl_2$

0.2% Gelrite

pH 5.7

EXAMPLE 15

ANALYSIS OF TRANSGENIC GLYCINE MAX PLANTS

While in the globular embryo state in liquid culture as described in Example 14, somatic soybean embryos contain very low amounts of triacylglycerol or storage proteins typical of maturing, zygotic soybean embryos. At this developmental stage, the ratio of total triacylglyceride to total polar lipid (phospholipids and glycolipid) is about 1:4, as is typical of zygotic soybean embryos at the developmental

stage from which the somatic embryo culture was initiated. At the globular stage as well, the mRNAs for the prominant seed proteins (a' subunit of β-conglycinin, Kunitz Trypsin Inhibitor III and Soybean Seed Lectin) are essentially absent. Upon transfer to hormone free media to allow differentiation to the maturing somatic embryo state as described in Example 14, triacylglycerol becomes the most abundant lipid class. As well, mRNAs for a'-subunit of β-conglycinin, Kunitz Trypsin Inhibitor III and Soybean Seed Lectin become very abundant messages in the total mRNA population. In these respects the somatic soybean embryo system behaves very similarly to maturing zygotic soybean embryos <u>in vivo</u>, and is therefore a good and rapid model system for analyzing the phenotypic effects of modifying the expression of genes in the fatty acid biosynthesis pathway. Similar somatic embryo culture systems have been documented and used in another oilseed crop, rapeseed (Taylor et al. (1990) Planta 181:18-26). Fatty acid analysis was performed as described in Example 12 using single embryos as the tissue source. A number of embryos from line 2872 (control tissue transformed with pCST) and lines 299,303,306 and 307 (line 2872 transformed with plasmid pCS3FdST1R) were analyzed for fatty acid content. The relative fatty-acid composition of embryos taken from tissue transformed with pCS3FdST1R was compared with control tissue, transformed with pCST. The results of this analysis are shown in Table 12.

<u>TABLE 12</u>

Line	Embryo	16:0	18:0	18:1	18:2	18:3
2872	1	17.7	4.1	11.3	52.8	14.1
	2	17.3	4.3	10.9	49.5	18.0
	3	16.1	4.1	13.8	48.2	17.3
	4	17.5	3.6	11.7	52.0	14.1

Line	Embryo	16:0	18:0	18:1	18:2	18:3
	5	16.6	3.9	12.7	53.7	12.6
	6	14.8	3.0	14.7	55.3	11.1
	av	16.7	3.8	12.5	51.9	14.5
299-1-3	1	16.5	4.1	9.7	61.4	6.3
299-15-1	1	14.7	3.6	11.9	61.3	8.4
	2	16.6	3.7	12.1	58.6	8.6
	3	16.7	4.1	14.9	53.2	11.1
	4	15.2	4.0	9.1	60.2	11.5
	5	16.0	4.2	13.9	55.2	10.7
	6	15.2	3.5	9.9	63.4	8.1
303-7-1	1	14.1	2.2	10.6	59.4	13.7
	2	14.0	2.8	12.5	59.3	11.4
306-4-5	1	17.5	4.2	8.1	62.7	7.4
	215.7	3.3	9.0	60.5	11.5	
	3	17.1	3.4	9.3	60.7	9.5
	4	15.7	3.8	9.2	61.2	9.7
	5	17.7	3.9	6.5	58.3	13.6
	6	16.6	3.4	10.2	59.2	10.6
306-4-8	1	16.6	3.9	15.3	50.7	11.8
	2	17.8	3.6	15.7	50.0	10.8
	3	16.7	3.3	11.1	52.0	14.6
	4	19.0	4.0	10.3	53.1	12.3
	5	19.7	3.5	9.0	53.6	13.0
	6	18.0	2.9	13.1	52.8	10.9
307-1-1	1	14.4	3.7	11.2	64.4	6.3
	2	15.4	3.4	7.8	61.0	11.3
	3	17.2	2.5	12.0	57.2	11.1
307-1-2	1	13.4	3.0	8.4	55.4	19.9
	2	16.3	3.1	6.4	55.7	18.7
	3	14.0	3.3	8.8	58.7	15.2
	4	15.8	2.5	9.8	59.7	12.2
	5	14.6	3.7	14.9	51.1	15.7
	6	14.3	3.9	11.4	55.5	14.1

Line	Embryo	16:0	18:0	18:1	18:2	18:3
307-1-3	1	14.8	3.1	9.4	60.5	12.2
	2	18.0	3.0	5.3	56.2	15.2
	3	18.0	3.4	2.5	58.6	15.4
307-1-4	1	15.0	2.7	13.8	61.7	6.9
	2	15.9	2.7	9.8	62.0	9.6
	3	14.6	3.2	13.4	61.4	6.7
307-1-5	1	15.9	3.5	7.6	61.7	11.2
	2	14.6	3.5	10.0	61.3	10.6
	3	18.7	2.6	6.8	53.0	19.0
307-1-7	1	15.3	3.5	12.5	60.3	8.5
	2	16.2	2.2	13.9	57.1	10.6
	3	14.9	3.1	12.2	58.0	11.8
307-1-9	1	16.4	2.9	23.2	47.9	9.6
	2	19.6	0.0	20.4	51.3	8.8
	3	16.8	3.3	24.6	49.6	5.7
307-1-11	1	18.1	3.6	5.7	52.9	19.7
	2	14.7	3.7	9.9	58.7	13.0
	3	15.1	3.7	11.3	55.8	14.1

The average 18:3 content of control embryos was 14.5% with a range from 11.1% to 18.0%. The average 18:3 content of transformed embryos was 11.5% with a range of 6.3% to 19.9%. Almost 80% of the transformed embryos (38/48) had an 18:3 content below that of the control mean. About 44% had an 18:3 content less than the lowest observed control value and 12.5% had an 18:3 content less than half of the control mean value (i.e., less than 7.5%). The lowest 18:3 content observed in transformed tissue was 6.3% (299-1-3, 307-1-2 #1) compared with the control low of 11.1%. In all cases in transformed tissue, a decrease in 18:3 content was reflected by an equivalent increase in 18:2 content indicating that the desaturation of 18:2 to 18:3 had

been reduced. The relative content of the the other fatty acids remained unchanged.

Southern analysis for the presence of the intact, introduced antisense construction was performed, as described in Example 12 using Bam HI cut gDNA, on a number of the transformed lines listed below using groups of embryos from a single transformation event. The approximate intact antisense copy number was estimated from the number and intensity of hybridizing bands on the autoradiograms and is shown in Table 13.

TABLE 13

Line No.	Antisense copy No.	18:3 (low)	18:3 (average)	18:2/18:3 ratio
2872	0	11.1	14.5	3.6
303-7/1	1	11.4	12.6	4.7
307-1/2	3	12.2	16.0	3.5
306-4/8	3	10.8	12.2	4.3
307-1/7	4	8.5	10.3	5.7
306-4/5	6	7.4	10.4	5.8
307-1/1	6	6.3	9.6	6.3
299-15/1	7	8.1	9.7	6.1
307-1/4	8	6.7	7.7	8.0

There was a reasonable correlation between intact antisense copy number and 18:3 content, an increase in copy number correlating with a decreased 18:3 content and a consequent increase in the 18:2/18:3 ratio. The average 18:2/18:3 ratio of line 307-1/4, which had at least 8 copies of the antisense cDNA, was more than twice that of the control.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANTS: Browse, John, Kinney, Anthony J., Pierce, John, Wierzbicki, Anna M., Yadav, Narendra S., Perez-Grau, Luis

(ii) TITLE OF INVENTION: Fatty Acid Desaturase Genes from Plants

(iii) NUMBER OF SEQUENCES: 32

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: E. I. du Pont de Nemours and Company
(B) STREET: 1007 Market Street
(C) CITY: Wilmington
(D) STATE: Delaware
(E) COUNTRY: U.S.A.
(F) ZIP: 19898

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: Macintosh
(C) OPERATING SYSTEM: Macintosh System, 6.0
(D) SOFTWARE: Microsoft Word, 4.0

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 07/804,259
(B) FILING DATE: 4 DECEMBER 1991

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Floyd, Linda A.
(B) REGISTRATION NUMBER: 33,692
(C) REFERENCE/DOCKET NUMBER: BB-1036-A

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (302) 992-4929
(B) TELEFAX: (302) 892-7949
(C) TELEX: 835420

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1350 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Arabidopsis thaliana IMMEDIATE SOURCE:
(B) CLONE: pCF3

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 46..1206

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CTCTCTCTCT CTCTCTTCTC TCTTTCTCTC CCCCTCTCTC CGGCG ATG GTT GTT       54
                                                  Met Val Val
                                                   1

GCT ATG GAC CAA CGC ACC AAT GTG AAC GGA GAT CCC GGC GCC GGA GAC    102
Ala Met Asp Gln Arg Thr Asn Val Asn Gly Asp Pro Gly Ala Gly Asp
      5                  10                  15

CGG AAG AAA GAA GAA AGG TTT GAT CCG AGT GCA CAA CCA CCG TTC AAG    150
Arg Lys Lys Glu Glu Arg Phe Asp Pro Ser Ala Gln Pro Pro Phe Lys
 20                  25                  30                  35

ATC GGA GAT ATA AGG GCG GCG ATT CCT AAG CAC TGT TGG GTT AAG AGT    198
Ile Gly Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp Val Lys Ser
                 40                  45                  50

CCT TTG AGA TCA ATG AGT TAC GTC GTC AGA GAC ATT ATC GCC GTC GCG    246
Pro Leu Arg Ser Met Ser Tyr Val Val Arg Asp Ile Ile Ala Val Ala
             55                  60                  65

GCT TTG GCC ATC GCT GCC GTG TAT GTT GAT AGC TGG TTC CTT TGG CCT    294
Ala Leu Ala Ile Ala Ala Val Tyr Val Asp Ser Trp Phe Leu Trp Pro
         70                  75                  80

CTT TAT TGG GCC GCC CAA GGA ACA CTT TTC TGG GCC ATC TTT GTT CTC    342
Leu Tyr Trp Ala Ala Gln Gly Thr Leu Phe Trp Ala Ile Phe Val Leu
     85                  90                  95

GGC CAC GAC TGT GGA CAT GGG AGT TTC TCA GAC ATT CCT CTA CTG AAT    390
Gly His Asp Cys Gly His Gly Ser Phe Ser Asp Ile Pro Leu Leu Asn
100                 105                 110                 115
```

```
AGT GTG GTT GGT CAC ATT CTT CAT TCT TTC ATC CTC GTT CCT TAC CAT    438
Ser Val Val Gly His Ile Leu His Ser Phe II  Leu Val Pro Tyr His
                120                 125                 130

GGT TGG AGA ATA AGC CAC CGG ACA CAC CAC CAG AAC CAT GGC CAT GTT    486
Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His Gly His Val
            135                 140                 145

GAA AAC GAC GAG TCA TGG GTT CCG TTA CCA GAA AGG GTG TAC AAG AAA    534
Glu Asn Asp Glu Ser Trp Val Pro Leu Pro Glu Arg Val Tyr Lys Lys
        150                 155                 160

TTG CCC CAC AGT ACT CGG ATG CTC AGA TAC ACT GTC CCT CTC CCC ATG    582
Leu Pro His Ser Thr Arg Met Leu Arg Tyr Thr Val Pro Leu Pro Met
    165                 170                 175

CTC GCA TAT CCT CTC TAT TTG TGC TAC AGA AGT CCT GGA AAA GAA GGA    630
Leu Ala Tyr Pro Leu Tyr Leu Cys Tyr Arg Ser Pro Gly Lys Glu Gly
180                 185                 190                 195

TCA CAT TTT AAC CCA TAC AGT AGT TTA TTT GCT CCA AGC GAG AGA AAG    678
Ser His Phe Asn Pro Tyr Ser Ser Leu Phe Ala Pro Ser Glu Arg Lys
                200                 205                 210

CTT ATT GCA ACT TCA ACT ACT TGT TGG TCC ATA ATG TTC GTC AGT CTT    726
Leu Ile Ala Thr Ser Thr Thr Cys Trp Ser Ile Met Phe Val Ser Leu
            215                 220                 225

ATC GCT CTA TCT TTC GTC TTC GGT CCA CTC GCG GTT CTT AAA GTC TAC    774
Ile Ala Leu Ser Phe Val Phe Gly Pro Leu Ala Val Leu Lys Val Tyr
        230                 235                 240

GGT GTA CCG TAC ATT ATC TTT GTG ATG TGG TTG GAT GCT GTC ACG TAT    822
Gly Val Pro Tyr Ile Ile Phe Val Met Trp Leu Asp Ala Val Thr Tyr
    245                 250                 255

TTG CAT CAT CAT GGT CAC GAT GAG AAG TTG CCT TGG TAT AGA GGC AAG    870
Leu His His His Gly His Asp Glu Lys Leu Pro Trp Tyr Arg Gly Lys
260                 265                 270                 275

GAA TGG AGT TAT CTA CGT GGA GGA TTA ACA ACA ATT GAT AGA GAT TAC    918
Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Ile Asp Arg Asp Tyr
                280                 285                 290

GGA ATC TTT AAC AAC ATT CAT CAC GAC ATT GGA ACT CAC GTG ATC CAT    966
Gly Ile Phe Asn Asn Ile His His Asp Ile Gly Thr His Val Ile His
            295                 300                 305

CAT CTC TTC CCA CAA ATC CCT CAC TAT CAC TTG GTC GAC GCC ACG AAA   1014
His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Asp Ala Thr Lys
        310                 315                 320

GCA GCT AAA CAT GTG TTG GGA AGA TAC TAC AGA GAA CCA AAG ACG TCA   1062
Ala Ala Lys His Val Leu Gly Arg Tyr Tyr Arg Glu Pro Lys Thr Ser
    325                 330                 335
```

GGA GCA ATA CCG ATC CAC TTG GTG GAG AGT TTG GTC GCA AGT ATT AAG 1110
Gly Ala Ile Pro Ile His Leu Val Glu Ser Leu Val Ala Ser Ile Lys
340 345 350 355

AAA GAT CAT TAC GTC AGC GAC ACT GGT GAT ATT GTC TTC TAC GAG ACA 1158
Lys Asp His Tyr Val Ser Asp Thr Gly Asp Ile Val Phe Tyr Glu Thr
360 365 370

GAT CCA GAT CTC TAC GTT TAC GCT TCT GAC AAA TCT AAA ATC AAT TAATCTCCAT 1213
Asp Pro Asp Leu Tyr Val Tyr Ala Ser Asp Lys Ser Lys Ile Asn
375 380 385

TTGTTTAGCT CTATTAGGAA TAAACCAGCC CACTTTTAAA ATTTTTATTT CTTGTTGTTT 1273

TTAAGTTAAA AGTGTACTCG TGAAACTCTT TTTTTTTTCT TTTTTTTTAT TAATGTATTT 1333

ACATTACAAG GCGTAAA 1350

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 386 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Val Val Ala Met Asp Gln Arg Thr Asn Val Asn Gly Asp Pro Gly
1 5 10 15

Ala Gly Asp Arg Lys Lys Glu Glu Arg Phe Asp Pro Ser Ala Gln Pro
20 25 30

Pro Phe Lys Ile Gly Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp
35 40 45

Val Lys Ser Pro Leu Arg Ser Met Ser Tyr Val Val Arg Asp Ile Ile
50 55 60

Ala Val Ala Ala Leu Ala Ile Ala Ala Val Tyr Val Asp Ser Trp Phe
65 70 75 80

Leu Trp Pro Leu Tyr Trp Ala Ala Gln Gly Thr Leu Phe Trp Ala Ile
85 90 95

Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe Ser Asp Ile Pro
100 105 110

Leu Leu Asn Ser Val Val Gly His Ile Leu His Ser Phe Ile Leu Val
115 120 125

Pro Tyr His Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His
130 135 140

Gly His Val Glu Asn Asp Glu Ser Trp Val Pro Leu Pro Glu Arg Val
145 150 155 160

Tyr Lys Lys Leu Pro His Ser Thr Arg Met Leu Arg Tyr Thr Val Pro
165 170 175

Leu Pro Met Leu Ala Tyr Pro Leu Tyr Leu Cys Tyr Arg Ser Pro Gly
180 185 190

Lys Glu Gly Ser His Phe Asn Pro Tyr Ser Ser Leu Phe Ala Pro Ser
195 200 205

Glu Arg Lys Leu Ile Ala Thr Ser Thr Thr Cys Trp Ser Ile Met Phe
210 215 220

Val Ser Leu Ile Ala Leu Ser Phe Val Phe Gly Pro Leu Ala Val Leu
225 230 235 240

Lys Val Tyr Gly Val Pro Tyr Ile Ile Phe Val Met Trp Leu Asp Ala
245 250 255

Val Thr Tyr Leu His His His Gly His Asp Glu Lys Leu Pro Trp Tyr
260 265 270

Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Ile Asp
275 280 285

Arg Asp Tyr Gly Ile Phe Asn Asn Ile His His Asp Ile Gly Thr His
290 295 300

Val Ile His His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Asp
305 310 315 320

Ala Thr Lys Ala Ala Lys His Val Leu Gly Arg Tyr Tyr Arg Glu Pro
325 330 335

Lys Thr Ser Gly Ala Ile Pro Ile His Leu Val Glu Ser Leu Val Ala
340 345 350

Ser Ile Lys Lys Asp His Tyr Val Ser Asp Thr Gly Asp Ile Val Phe
355 360 365

Tyr Glu Thr Asp Pro Asp Leu Tyr Val Tyr Ala Ser Asp Lys Ser Lys
370 375 380

Ile Asn
385

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 255 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (g nomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Arabidopsis thaliana

(vii) IMMEDIATE SOURCE:

(B) CLONE: pF1

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 68..255

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
AAATTCATCA AACCCTTTCT TCACCACATT ATTTTCACTG AGCGCATAAC ATTTTTGAGA     60
CAAGAGACTC TCTCTCTCTC TCTTCTCTCT TTCTCTCCCC CTCTCTCCGG CGATGGTTGT    120
TGCTATGGAC CAACGCACCA ATGTGAACGG AGATCCCGGC GCCGGAGACC GGAAGAAAGA    180
AGAAAGGTTT GATCCGAGTG CACAACCACC GTTCAAGATC GGAGATATAA GGGCGGCGAT    240
TCCTAAGCAC TGTTG                                                     255
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1525 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Arabidopsis thaliana

(vii) IMMEDIATE SOURCE:

(B) CLONE: pACF2-2

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 10..1350

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
CAAGTTCTA ATG GCG AAC TTG GTC TTA TCA GAA TGT GGT ATA CGA CCT     48
          Met Ala Asn Leu Val Leu Ser Glu Cys Gly Ile Arg Pro
            1               5                  10

CTC CCC AGA ATC TAC ACA ACA CCC AGA TCC AAT TTC CTC TCC AAC AAC     96
Leu Pro Arg Ile Tyr Thr Thr Pro Arg Ser Asn Phe Leu Ser Asn Asn
     15                  20                  25

AAC AAA TTC AGA CCA TCA CTT TCT TCT TCT TCT TAC AAA ACA TCA TCA    144
Asn Lys Phe Arg Pro Ser Leu Ser Ser Ser Ser Tyr Lys Thr Ser Ser
 30                  35                  40                  45

TCT CCT CTG TCT TTT GGT CTG AAT TCA CGA GAT GGG TTC ACG AGG AAT    192
Ser Pro Leu Ser Phe Gly Leu Asn Ser Arg Asp Gly Phe Thr Arg Asn
                 50                  55                  60

TGG GCG TTG AAT GTG AGC ACA CCA TTA ACG ACA CCA ATA TTT GAG GAG    240
Trp Ala Leu Asn Val Ser Thr Pro Leu Thr Thr Pro Ile Phe Glu Glu
             65                  70                  75

TCT CCA TTG GAG GAA GAT AAT AAA CAG AGA TTC GAT CCA GGT GCG CCT    288
Ser Pro Leu Glu Glu Asp Asn Lys Gln Arg Phe Asp Pro Gly Ala Pro
         80                  85                  90

CCT CCG TTC AAT TTA GCT GAT ATT AGA GCA GCT ATA CCT AAG CAT TGT    336
Pro Pro Phe Asn Leu Ala Asp Ile Arg Ala Ala Ile Pro Lys His Cys
     95                 100                 105

TGG GTT AAG AAT CCA TGG AAG TCT TTG AGT TAT GTC GTC AGA GAC GTC    384
Trp Val Lys Asn Pro Trp Lys Ser Leu Ser Tyr Val Val Arg Asp Val
110                 115                 120                 125

GCT ATC GTC TTT GCA TTG GCT GCT GGA GCT GCT TAC CTC AAC AAT TGG    432
Ala Ile Val Phe Ala Leu Ala Ala Gly Ala Ala Tyr Leu Asn Asn Trp
                130                 135                 140

ATT GTT TGG CCT CTC TAT TGG CTC GCT CAA GGA ACC ATG TTT TGG GCT    480
Ile Val Trp Pro Leu Tyr Trp Leu Ala Gln Gly Thr Met Phe Trp Ala
            145                 150                 155

CTC TTT GTT CTT GGT CAT GAC TGT GGA CAT GGT AGT TTC TCA AAT GAT    528
Leu Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe Ser Asn Asp
        160                 165                 170

CCG AAG TTG AAC AGT GTG GTC GGT CAT CTT CTT CAT TCC TCA ATT CTG    576
Pro Lys Leu Asn Ser Val Val Gly His Leu Leu His Ser Ser Ile Leu
    175                 180                 185

GTC CCA TAC CAT GGC TGG AGA ATT AGT CAC AGA ACT CAC CAC CAG AAC    624
Val Pro Tyr His Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn
190                 195                 200                 205

CAT GGA CAT GTT GAG AAT GAC GAA TCT TGG CAT CCT ATG TCT GAG AAA    672
His Gly His Val Glu Asn Asp Glu Ser Trp His Pro Met Ser Glu Lys
                210                 215                 220
```

```
ATC TAC AAT ACT TTG GAC AAG CCG ACT AGA TTC TTT AGA TTT ACA CTG      720
Ile Tyr Asn Thr Leu Asp Lys Pro Thr Arg Ph  Phe Arg Phe Thr Leu
            225                 230                 235

CCT CTC GTG ATG CTT GCA TAC CCT TTC TAC TTG TGG GCT CGA AGT CCG      768
Pro Leu Val Met Leu Ala Tyr Pro Phe Tyr Leu Trp Ala Arg Ser Pro
        240                 245                 250

GGG AAA AAG GGT TCT CAT TAC CAT CCA GAC AGT GAC TTG TTC CTC CCT      816
Gly Lys Lys Gly Ser His Tyr His Pro Asp Ser Asp Leu Phe Leu Pro
    255                 260                 265

AAA GAG AGA AAG GAT GTC CTC ACT TCT ACT GCT TGT TGG ACT GCA ATG      864
Lys Glu Arg Lys Asp Val Leu Thr Ser Thr Ala Cys Trp Thr Ala Met
270                 275                 280                 285

GCT GCT CTG CTT GTT TGT CTC AAC TTC ACA ATC GGT CCA ATT CAA ATG      912
Ala Ala Leu Leu Val Cys Leu Asn Phe Thr Ile Gly Pro Ile Gln Met
                290                 295                 300

CTC AAA CTT TAT GGA ATT CCT TAC TGG ATA AAT GTA ATG TGG TTG GAC      960
Leu Lys Leu Tyr Gly Ile Pro Tyr Trp Ile Asn Val Met Trp Leu Asp
            305                 310                 315

TTT GTG ACT TAC CTG CAT CAC CAT GGT CAT GAA GAT AAG CTT CCT TGG     1008
Phe Val Thr Tyr Leu His His His Gly His Glu Asp Lys Leu Pro Trp
        320                 325                 330

TAC CGT GGC AAG GAG TGG AGT TAC CTG AGA GGA GGA CTT ACA ACA TTG     1056
Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Leu
    335                 340                 345

GAT CGT GAC TAC GGA TTG ATC AAT AAC ATC CAT CAT GAT ATT GGA ACT     1104
Asp Arg Asp Tyr Gly Leu Ile Asn Asn Ile His His Asp Ile Gly Thr
350                 355                 360                 365

CAT GTG ATA CAT CAT CTT TTC CCG CAG ATC CCA CAT TAT CAT CTA GTA     1152
His Val Ile His His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val
                370                 375                 380

GAA GCA ACA GAA GCA GCT AAA CCA GTA TTA GGG AAG TAT TAC AGG GAG     1200
Glu Ala Thr Glu Ala Ala Lys Pro Val Leu Gly Lys Tyr Tyr Arg Glu
            385                 390                 395

CCT GAT AAG TCT GGA CCG TTG CCA TTA CAT TTA CTG GAA ATT CTA GCG     1248
Pro Asp Lys Ser Gly Pro Leu Pro Leu His Leu Leu Glu Ile Leu Al
        400                 405                 410

AAA AGT ATA AAA GAA GAT CAT TAC GTG AGC GAC GAA GGA GAA GTT GTA     1296
Lys Ser Ile Lys Glu Asp His Tyr Val Ser Asp Glu Gly Glu Val Val
    415                 420                 425

TAC TAT AAA GCA GAT CCA AAT CTC TAT GGA GAG GTC AAA GTA AGA GCA     1344
Tyr Tyr Lys Ala Asp Pro Asn Leu Tyr Gly Glu Val Lys Val Arg Ala
430                 435                 440                 445

GAT TGAAATGAAG CAGGCTTGAG ATTGAAGTTT TTTCTATTTC AGACCAGCTG          1397
Asp
```

ATTTTTTGCT TACTGTATCA ATTTATTGTG TCACCCACCA GAGAGTTAGT ATCTCTGAAT 1457

ACGATCGATC AGATGGAAAC AACAAATTTG TTTGCGATAC TGAAGCTATA TATACCATAC 1517

ATTGCATT 1525

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 446 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Met Ala Asn Leu Val Leu Ser Glu Cys Gly Ile Arg Pro Leu Pro Arg
 1               5                  10                  15

Ile Tyr Thr Thr Pro Arg Ser Asn Phe Leu Ser Asn Asn Asn Lys Phe
            20                  25                  30

Arg Pro Ser Leu Ser Ser Ser Ser Tyr Lys Thr Ser Ser Ser Pro Leu
        35                  40                  45

Ser Phe Gly Leu Asn Ser Arg Asp Gly Phe Thr Arg Asn Trp Ala Leu
    50                  55                  60

Asn Val Ser Thr Pro Leu Thr Thr Pro Ile Phe Glu Glu Ser Pro Leu
65                  70                  75                  80

Glu Glu Asp Asn Lys Gln Arg Phe Asp Pro Gly Ala Pro Pro Pro Phe
                85                  90                  95

Asn Leu Ala Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp Val Lys
            100                 105                 110

Asn Pro Trp Lys Ser Leu Ser Tyr Val Val Arg Asp Val Ala Ile Val
        115                 120                 125

Phe Ala Leu Ala Ala Gly Ala Ala Tyr Leu Asn Asn Trp Ile Val Trp
    130                 135                 140

Pro Leu Tyr Trp Leu Ala Gln Gly Thr Met Phe Trp Ala Leu Phe Val
145                 150                 155                 160

Leu Gly His Asp Cys Gly His Gly Ser Phe Ser Asn Asp Pro Lys Leu
                165                 170                 175

Asn Ser Val Val Gly His Leu Leu His Ser Ser Ile Leu Val Pro Tyr
            180                 185                 190

His Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His Gly His
        195                 200                 205
```

Val Glu Asn Asp Glu Ser Trp His Pro Met Ser Glu Lys Il Tyr Asn
210 215 220

Thr Leu Asp Lys Pro Thr Arg Phe Phe Arg Phe Thr Leu Pro Leu Val
225 230 235 240

Met Leu Ala Tyr Pro Phe Tyr Leu Trp Ala Arg Ser Pro Gly Lys Lys
245 250 255

Gly Ser His Tyr His Pro Asp Ser Asp Leu Phe Leu Pro Lys Glu Arg
260 265 270

Lys Asp Val Leu Thr Ser Thr Ala Cys Trp Thr Ala Met Ala Ala Leu
275 280 285

Leu Val Cys Leu Asn Phe Thr Ile Gly Pro Ile Gln Met Leu Lys Leu
290 295 300

Tyr Gly Ile Pro Tyr Trp Ile Asn Val Met Trp Leu Asp Phe Val Thr
305 310 315 320

Tyr Leu His His His Gly His Glu Asp Lys Leu Pro Trp Tyr Arg Gly
325 330 335

Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Leu Asp Arg Asp
340 345 350

Tyr Gly Leu Ile Asn Asn Ile His His Asp Ile Gly Thr His Val Ile
355 360 365

His His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Glu Ala Thr
370 375 380

Glu Ala Ala Lys Pro Val Leu Gly Lys Tyr Tyr Arg Glu Pro Asp Lys
385 390 395 400

Ser Gly Pro Leu Pro Leu His Leu Leu Glu Ile Leu Ala Lys Ser Ile
405 410 415

Lys Glu Asp His Tyr Val Ser Asp Glu Gly Glu Val Val Tyr Tyr Lys
420 425 430

Ala Asp Pro Asn Leu Tyr Gly Glu Val Lys Val Arg Ala Asp
435 440 445

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1429 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Brassica napus

(vii) IMMEDIATE SOURCE:

(B) CLONE: pBNSF3-f2

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 79..1212

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
TTCAAATTCA GACAATCCCC TTCTTCTCCC CGGTTTCGTC TGAACTCTCG AAACTGGGCG       60

TTGAATGTAA CCACACCT CTA ACA GTC GAC TCC TCA TCA TCT CCT CCA ATC       111
                    Leu Thr Val Asp Ser Ser Ser Ser Pro Pro Ile
                     1               5                  10

GAG GAA GAA CCC AAA ACG CAG AGA TTC GAC CCA GGC GCT CCT CCT CCG       159
Glu Glu Glu Pro Lys Thr Gln Arg Phe Asp Pro Gly Ala Pro Pro Pro
             15                  20                  25

TTC AAC CTA GCT GAC ATC AGA GCG GCG ATA CCT AAG CAT TGC TGG GTT       207
Phe Asn Leu Ala Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp Val
         30                  35                  40

AAG AAT CCA TGG AAG TCT ATG AGT TAC GTC GTC AGA GAG CTA GCC ATC       255
Lys Asn Pro Trp Lys Ser Met Ser Tyr Val Val Arg Glu Leu Ala Ile
     45                  50                  55

GTG TTC GCA CTA GCT GCT GGA GCT GCT TAC CTC AAC AAT TGG CTT GTT       303
Val Phe Ala Leu Ala Ala Gly Ala Ala Tyr Leu Asn Asn Trp Leu Val
 60                  65                  70                  75

TGG CCT CTC TAT TGG ATT GCT CAA GGA ACC ATG TTC TGG GCT CTC TTT       351
Trp Pro Leu Tyr Trp Ile Ala Gln Gly Thr Met Phe Trp Ala Leu Phe
                 80                  85                  90

GTT CTT GGC CAT GAC TGT GGA CAT GGA AGC TTC TCA AAT GAT CCG AGG       399
Val Leu Gly His Asp Cys Gly His Gly Ser Phe Ser Asn Asp Pro Arg
             95                 100                 105

TTG AAC AGT GTG GTG GGT CAC CTT CTT CAT TCC TCT ATT CTA GTC CCT       447
Leu Asn Ser Val Val Gly His Leu Leu His Ser Ser Ile Leu Val Pro
        110                 115                 120

TAC CAT GGC TGG AGA ATT AGC CAC AGA ACT CAC CAC CAG AAC CAT GGA       495
Tyr His Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His Gly
    125                 130                 135

CAT GTT GAG AAC GAT GAA TCT TGG CAT CCT ATG TCT GAG AAA ATC TAC       543
His Val Glu Asn Asp Glu Ser Trp His Pro Met Ser Glu Lys Ile Tyr
140                 145                 150                 155
```

```
AAG AGT TTG GAC AAA CCC ACT CGG TTC TTT AGA TTT ACA TTG CCT CTC      591
Lys Ser Leu Asp Lys Pro Thr Arg Phe Phe Arg Phe Thr Leu Pro Leu
                160                 165                 170

GTG ATG CTC GCT TAC CCT TTC TAC TTG TGG GCA AGA AGT CCA GGG AAG      639
Val Met Leu Ala Tyr Pro Phe Tyr Leu Trp Ala Arg Ser Pro Gly Lys
            175                 180                 185

AAG GGT TCT CAT TAC CAT CCA GAC AGC GAC TTG TTC CTT CCT AAA GAG      687
Lys Gly Ser His Tyr His Pro Asp Ser Asp Leu Phe Leu Pro Lys Glu
        190                 195                 200

AGA AAC GAT GTT CTC ACT TCT ACC GCT TGT TGG ACT GCA ATG GCT GTT      735
Arg Asn Asp Val Leu Thr Ser Thr Ala Cys Trp Thr Ala Met Ala Val
    205                 210                 215

CTG CTT GTC TGT CTC AAC TTC GTG ATG GGT CCA ATG CAA ATG CTC AAA      783
Leu Leu Val Cys Leu Asn Phe Val Met Gly Pro Met Gln Met Leu Lys
220                 225                 230                 235

CTT TAT GTC ATT CCT TAC TGG ATA AAT GTA ATG TGG TTG GAC TTT GTG      831
Leu Tyr Val Ile Pro Tyr Trp Ile Asn Val Met Trp Leu Asp Phe Val
                240                 245                 250

ACT TAC CTG CAT CAC CAT GGT CAT GAA GAT AAG CTC CCT TGG TAC CGT      879
Thr Tyr Leu His His His Gly His Glu Asp Lys Leu Pro Trp Tyr Arg
            255                 260                 265

GGG AAG GAA TGG AGT TAC TTG AGA GGA GGA CTT ACA ACA TTG GAC CGG      927
Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Leu Asp Arg
        270                 275                 280

GAC TAC GGA TTG ATC AAC AAC ATC CAT CAC GAC ATT GGA ACT CAT GTG      975
Asp Tyr Gly Leu Ile Asn Asn Ile His His Asp Ile Gly Thr His Val
    285                 290                 295

ATA CAT CAT CTT TTC CCT CAG ATC CCA CAT TAT CAT CTA GTA GAA GCA     1023
Ile His His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Glu Ala
300                 305                 310                 315

ACA GAA GCA GCT AAA CCA GTA TTA GGG AAG TAT TAT AGG GAG CCT GAT     1071
Thr Glu Ala Ala Lys Pro Val Leu Gly Lys Tyr Tyr Arg Glu Pro Asp
                320                 325                 330

AAG TCT GGA CCT TTG CCA TTA CAT TTA CTG GGA ATC TTA GCA AAA AGT     1119
Lys Ser Gly Pro Leu Pro Leu His Leu Leu Gly Ile Leu Ala Lys Ser
            335                 340                 345

ATT AAA GAA GAT CAT TTT GTG AGC GAT GAA GGA GAT GTT GTA TAC TAT     1167
Ile Lys Glu Asp His Phe Val Ser Asp Glu Gly Asp Val Val Tyr Tyr
        350                 355                 360

GAA GCA GAC CCT AAT CTC TAT GGA GAG ATC AAG GTA ACA GCA GAG         1212
Glu Ala Asp Pro Asn Leu Tyr Gly Glu Ile Lys Val Thr Ala Glu
    365                 370                 375

TGAAATGAAG CTGTCAGATT TATCTATTTC TGACCAGCTG ATTTTTTTTG CTTATTAATG    1272
```

TCAATTCATT GTGTTACCAT TATCTCTGAA TACAATCAGA TGGAAACCCC AACTTTGTTT 1332

TCAATACTTG AAGCTATATA TATATATATA TATGTAAGAT ACATTGTATT GTCATTAGAT 1392

TCACCATTCT CAAGGTTCTT ATACAAAAAA AAAAAAA 1429

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 378 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Leu Thr Val Asp Ser Ser Ser Ser Pro Pro Ile Glu Glu Glu Pro Lys
1 5 10 15

Thr Gln Arg Phe Asp Pro Gly Ala Pro Pro Pro Phe Asn Leu Ala Asp
20 25 30

Ile Arg Ala Ala Ile Pro Lys His Cys Trp Val Lys Asn Pro Trp Lys
35 40 45

Ser Met Ser Tyr Val Val Arg Glu Leu Ala Ile Val Phe Ala Leu Ala
50 55 60

Ala Gly Ala Ala Tyr Leu Asn Asn Trp Leu Val Trp Pro Leu Tyr Trp
65 70 75 80

Ile Ala Gln Gly Thr Met Phe Trp Ala Leu Phe Val Leu Gly His Asp
85 90 95

Cys Gly His Gly Ser Phe Ser Asn Asp Pro Arg Leu Asn Ser Val Val
100 105 110

Gly His Leu Leu His Ser Ser Ile Leu Val Pro Tyr His Gly Trp Arg
115 120 125

Ile Ser His Arg Thr His His Gln Asn His Gly His Val Glu Asn Asp
130 135 140

Glu Ser Trp His Pro Met Ser Glu Lys Ile Tyr Lys Ser Leu Asp Lys
145 150 155 160

Pro Thr Arg Phe Phe Arg Phe Thr Leu Pro Leu Val Met Leu Ala Tyr
165 170 175

Pro Phe Tyr Leu Trp Ala Arg Ser Pro Gly Lys Lys Gly Ser His Tyr
180 185 190

His Pro Asp Ser Asp Leu Phe Leu Pro Lys Glu Arg Asn Asp Val Leu
195 200 205

```
Thr Ser Thr Ala Cys Trp Thr Ala Met Ala Val Leu Leu Val Cys Leu
    210                 215                 220

Asn Phe Val Met Gly Pro Met Gln Met Leu Lys Leu Tyr Val Ile Pro
225                 230                 235                 240

Tyr Trp Ile Asn Val Met Trp Leu Asp Phe Val Thr Tyr Leu His His
                245                 250                 255

His Gly His Glu Asp Lys Leu Pro Trp Tyr Arg Gly Lys Glu Trp Ser
            260                 265                 270

Tyr Leu Arg Gly Gly Leu Thr Thr Leu Asp Arg Asp Tyr Gly Leu Ile
        275                 280                 285

Asn Asn Ile His His Asp Ile Gly Thr His Val Ile His His Leu Phe
    290                 295                 300

Pro Gln Ile Pro His Tyr His Leu Val Glu Ala Thr Glu Ala Ala Lys
305                 310                 315                 320

Pro Val Leu Gly Lys Tyr Tyr Arg Glu Pro Asp Lys Ser Gly Pro Leu
                325                 330                 335

Pro Leu His Leu Leu Gly Ile Leu Ala Lys Ser Ile Lys Glu Asp His
            340                 345                 350

Phe Val Ser Asp Glu Gly Asp Val Val Tyr Tyr Glu Ala Asp Pro Asn
        355                 360                 365

Leu Tyr Gly Glu Ile Lys Val Thr Ala Glu
    370                 375
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1429 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Brassica napus

(vii) IMMEDIATE SOURCE:

(B) CLONE: pBNSFd-2

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1..1215

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
TTC AAA TTC AGA CAA TCC CCT TCT TCT CCC CGG TTT CGT CTG AAC TCT      48
Phe Lys Phe Arg Gln Ser Pro Ser Ser Pro Arg Phe Arg Leu Asn Ser
  1               5                  10                  15

CGA AAC TGG GCG TTG AAT GTA ACC ACA CCT CTA ACA GTC GAC TCC TCA      96
Arg Asn Trp Ala Leu Asn Val Thr Thr Pro Leu Thr Val Asp Ser Ser
             20                  25                  30

TCA TCT CCT CCA ATC GAG GAA GAA CCC AAA ACG CAG AGA TTC GAC CCA     144
Ser Ser Pro Pro Ile Glu Glu Glu Pro Lys Thr Gln Arg Phe Asp Pro
         35                  40                  45

GGC GCT CCT CCT CCG TTC AAC CTA GCT GAC ATC AGA GCG GCG ATA CCT     192
Gly Ala Pro Pro Pro Phe Asn Leu Ala Asp Ile Arg Ala Ala Ile Pro
     50                  55                  60

AAG CAT TGC TGG GTT AAG AAT CCA TGG AAG TCT ATG AGT TAC GTC GTC     240
Lys His Cys Trp Val Lys Asn Pro Trp Lys Ser Met Ser Tyr Val Val
 65                  70                  75                  80

AGA GAG CTA GCC ATC GTG TTC GCA CTA GCT GCT GGA GCT GCT TAC CTC     288
Arg Glu Leu Ala Ile Val Phe Ala Leu Ala Ala Gly Ala Ala Tyr Leu
                 85                  90                  95

AAC AAT TGG CTT GTT TGG CCT CTC TAT TGG ATT GCT CAA GGA ACC ATG     336
Asn Asn Trp Leu Val Trp Pro Leu Tyr Trp Ile Ala Gln Gly Thr Met
            100                 105                 110

TTC TGG GCT CTC TTT GTT CTT GGC CAT GAC TGT GGA CAT GGA AGC TTC     384
Phe Trp Ala Leu Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe
        115                 120                 125

TCA AAT GAT CCG AGG TTG AAC AGT GTG GTG GGT CAC CTT CTT CAT TCC     432
Ser Asn Asp Pro Arg Leu Asn Ser Val Val Gly His Leu Leu His Ser
    130                 135                 140

TCT ATT CTA GTC CCT TAC CAT GGC TGG AGA ATT AGC CAC AGA ACT CAC     480
Ser Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His Arg Thr His
145                 150                 155                 160

CAC CAG AAC CAT GGA CAT GTT GAG AAC GAT GAA TCT TGG CAT CCT ATG     528
His Gln Asn His Gly His Val Glu Asn Asp Glu Ser Trp His Pro Met
                165                 170                 175

TCT GAG AAA ATC TAC AAG AGT TTG GAC AAA CCC ACT CGG TTC TTT AGA     576
Ser Glu Lys Ile Tyr Lys Ser Leu Asp Lys Pro Thr Arg Phe Phe Arg
            180                 185                 190

TTT ACA TTG CCT CTC GTG ATG CTC GCT TAC CCT TTC TAC TTG TGG GCA     624
Phe Thr Leu Pro Leu Val Met Leu Ala Tyr Pro Phe Tyr Leu Trp Ala
        195                 200                 205
```

```
AGA AGT CCA GGG AAG AAG GGT TCT CAT TAC CAT CCA GAC AGC GAC TTG      672
Arg Ser Pro Gly Lys Lys Gly Ser His Tyr His Pro Asp Ser Asp Leu
    210                 215                 220

TTC CTT CCT AAA GAG AGA AAC GAT GTT CTC ACT TCT ACC GCT TGT TGG      720
Phe Leu Pro Lys Glu Arg Asn Asp Val Leu Thr Ser Thr Ala Cys Trp
225                 230                 235                 240

ACT GCA ATG GCT GTT CTG CTT GTC TGT CTC AAC TTC GTG ATG GGT CCA      768
Thr Ala Met Ala Val Leu Leu Val Cys Leu Asn Phe Val Met Gly Pro
                245                 250                 255

ATG CAA ATG CTC AAA CTT TAT GTC ATT CCT TAC TGG ATA AAT GTA ATG      816
Met Gln Met Leu Lys Leu Tyr Val Ile Pro Tyr Trp Ile Asn Val Met
            260                 265                 270

TGG TTG GAC TTT GTG ACT TAC CTG CAT CAC CAT GGT CAT GAA GAT AAG      864
Trp Leu Asp Phe Val Thr Tyr Leu His His His Gly His Glu Asp Lys
        275                 280                 285

CTC CCT TGG TAC CGT GGG AAG GAA TGG AGT TAC TTG AGA GGA GGA CTT      912
Leu Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu
    290                 295                 300

ACA ACA TTG GAC CGG GAC TAC GGA TTG ATC AAC AAC ATC CAT CAC GAC      960
Thr Thr Leu Asp Arg Asp Tyr Gly Leu Ile Asn Asn Ile His His Asp
305                 310                 315                 320

ATT GGA ACT CAT GTG ATA CAT CAT CTT TTC CCT CAG ATC CCA CAT TAT     1008
Ile Gly Thr His Val Ile His His Leu Phe Pro Gln Ile Pro His Tyr
                325                 330                 335

CAT CTA GTA GAA GCA ACA GAA GCA GCT AAA CCA GTA TTA GGG AAG TAT     1056
His Leu Val Glu Ala Thr Glu Ala Ala Lys Pro Val Leu Gly Lys Tyr
            340                 345                 350

TAT AGG GAG CCT GAT AAG TCT GGA CCT TTG CCA TTA CAT TTA CTG GGA     1104
Tyr Arg Glu Pro Asp Lys Ser Gly Pro Leu Pro Leu His Leu Leu Gly
        355                 360                 365

ATC TTA GCA AAA AGT ATT AAA GAA GAT CAT TTT GTG AGC GAT GAA GGA     1152
Ile Leu Ala Lys Ser Ile Lys Glu Asp His Phe Val Ser Asp Glu Gly
    370                 375                 380

GAT GTT GTA TAC TAT GAA GCA GAC CCT AAT CTC TAT GGA GAG ATC AAG     1200
Asp Val Val Tyr Tyr Glu Ala Asp Pro Asn Leu Tyr Gly Glu Ile Lys
385                 390                 395                 400

GTA ACA GCA GAG TGAAATGAAG CTGTCAGATT TATCTATTTC TGACCAGCTG         1252
Val Thr Ala Glu
                405

ATTTTTTTTG CTTATTAATG TCAATTCATT GTGTTACCAT TATCTCTGAA TACAATCAGA   1312

TGGAAACCCC AACTTTGTTT TCAATACTTG AAGCTATATA TATATATATA TATGTAAGAT   1372

ACATTGTATT GTCATTAGAT TCACCATTCT CAAGGTTCTT ATACAAAAAA AAAAAAA      1429
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 404 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Phe Lys Phe Arg Gln Ser Pro Ser Ser Pro Arg Phe Arg Leu Asn Ser
1 5 10 15

Arg Asn Trp Ala Leu Asn Val Thr Thr Pro Leu Thr Val Asp Ser Ser
20 25 30

Ser Ser Pro Pro Ile Glu Glu Glu Pro Lys Thr Gln Arg Phe Asp Pro
35 40 45

Gly Ala Pro Pro Pro Phe Asn Leu Ala Asp Ile Arg Ala Ala Ile Pro
50 55 60

Lys His Cys Trp Val Lys Asn Pro Trp Lys Ser Met Ser Tyr Val Val
65 70 75 80

Arg Glu Leu Ala Ile Val Phe Ala Leu Ala Ala Gly Ala Ala Tyr Leu
85 90 95

Asn Asn Trp Leu Val Trp Pro Leu Tyr Trp Ile Ala Gln Gly Thr Met
100 105 110

Phe Trp Ala Leu Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe
115 120 125

Ser Asn Asp Pro Arg Leu Asn Ser Val Val Gly His Leu Leu His Ser
130 135 140

Ser Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His Arg Thr His
145 150 155 160

His Gln Asn His Gly His Val Glu Asn Asp Glu Ser Trp His Pro Met
165 170 175

Ser Glu Lys Ile Tyr Lys Ser Leu Asp Lys Pro Thr Arg Phe Phe Arg
180 185 190

Phe Thr Leu Pro Leu Val Met Leu Ala Tyr Pro Phe Tyr Leu Trp Ala
195 200 205

Arg Ser Pro Gly Lys Lys Gly Ser His Tyr His Pro Asp Ser Asp Leu
210 215 220

Phe Leu Pro Lys Glu Arg Asn Asp Val Leu Thr Ser Thr Ala Cys Trp
225 230 235 240

Thr Ala Met Ala Val Leu Leu Val Cys Leu Asn Phe Val Met Gly Pro
245 250 255

Met Gln Met Leu Lys Leu Tyr Val Ile Pro Tyr Trp Ile Asn Val Met
260 265 270

Trp Leu Asp Phe Val Thr Tyr Leu His His His Gly His Glu Asp Lys
275 280 285

Leu Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu
290 295 300

Thr Thr Leu Asp Arg Asp Tyr Gly Leu Ile Asn Asn Ile His His Asp
305 310 315 320

Ile Gly Thr His Val Ile His His Leu Phe Pro Gln Ile Pro His Tyr
325 330 335

His Leu Val Glu Ala Thr Glu Ala Ala Lys Pro Val Leu Gly Lys Tyr
340 345 350

Tyr Arg Glu Pro Asp Lys Ser Gly Pro Leu Pro Leu His Leu Leu Gly
355 360 365

Ile Leu Ala Lys Ser Ile Lys Glu Asp His Phe Val Ser Asp Glu Gly
370 375 380

Asp Val Val Tyr Tyr Glu Ala Asp Pro Asn Leu Tyr Gly Glu Ile Lys
385 390 395 400

Val Thr Ala Glu

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2181 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Glycine max

(vii) IMMEDIATE SOURCE:

(B) CLONE: pXF1

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 855..1997

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
ACAATAATAA ATCCATATTT TTATAATTAA AAGTAGTAGA TTACAGCGAT GCACTTGAGA      60

AACATATTAA GTGGACTAAT TCTCCCTGGT CAAGCAAGAA AAAAACCAGC TATGACCCAA     120

GGTAGAGAGA GATTATACAC AGAATACTAG TAATTAACTA AGACTGGCTC TGCAATTGCC     180

AAAAACTCCA TTGCAGTAGC AGCCACCTGA GAAGACACTA AGACCTAGAC TAGACCATAC     240

ATATGAAGAT TAATACGCTT ACATAACAAC ATAGGACACT AAGAAAACAC GGCTTACAGA     300

GAATCCAGCT GACTCTATAA GAGGGGTACT TCTGGAGATT AAAATTATCC GAATCACCTT     360

CCCACTGCGG CTGCTGACGT CAGCGAAAGT CAGAACCGAA AGCGGCGAAG AACCTTCAGA     420

AGAGGAGGAA GCACTTCGAC CTTACAAGAG TTGTTGTCGT TGTTGTTGTC GTTCTCTGGC     480

GGAGAAGCGA GTTTGGATCG CGTTTTCCTC GGAGGCTTCT CGGTCTTCCC CTGTTTCTGC     540

AGCTCAGCCA GGCCCTCGCA AATGGCCTGA AGCTTGGCGT CAACGGCGGA ATGAAGAGGC     600

TAATACTCCC CGAAGTCACC ACCGACGGAG GAACCCTGGT GTCGGAGGTT GGGGAAGTTG     660

AGCCTGGCGA AGTCACCTCG GAGCTTGTAC GCGGCCTTGT GGTACGCCAG AGCGGCTTCC     720

TCGGCGGTGT CGAAGGTTCC CAGCCATAGC CTGGTCCGGA TTCTTCGGGA GTCTAATCTC     780

AGCCACCCAC TTCCCCCCTG AGAAAAGAGA GGAACCACAC TCTCTAAGCC AAAGCAAAAG     840

CAGCAGCAGC AGCA ATG GTT AAA GAC ACA AAG CCT TTA GCC TAT GCT GCC      890
                Met Val Lys Asp Thr Lys Pro Leu Ala Tyr Ala Ala
                 1               5                  10

AAT AAT GGA TAC CAA CAA AAG GGT TCT TCT TTT GAT TTT GAT CCT AGC      938
Asn Asn Gly Tyr Gln Gln Lys Gly Ser Ser Phe Asp Phe Asp Pro Ser
         15                  20                  25

GCT CCT CCA CCG TTT AAG ATT GCA GAA ATC AGA GCT TCA ATA CCA AAA      986
Ala Pro Pro Pro Phe Lys Ile Ala Glu Ile Arg Ala Ser Ile Pro Lys
    30                  35                  40

CAT TGC TGG GTC AAG AAT CCA TGG AGA TCC CTC AGT TAT GTT CTC AGG     1034
His Cys Trp Val Lys Asn Pro Trp Arg Ser Leu Ser Tyr Val Leu Arg
45                  50                  55                  60

GAT GTG CTT GTA ATT GCT GCA TTG GTG GCT GCA GCA ATT CAC TTC GAC     1082
Asp Val Leu Val Ile Ala Ala Leu Val Ala Ala Ala Ile His Phe Asp
                65                  70                  75

AAC TGG CTT CTC TGG CTA ATC TAT TGC CCC ATT CAA GGC ACA ATG TTC     1130
Asn Trp Leu Leu Trp Leu Ile Tyr Cys Pro Ile Gln Gly Thr Met Phe
            80                  85                  90
```

TGG GCT CTC TTT GTT CTT GGA CAT GAT TGT GGC CAT GGA AGC TTT TCA 1178
Trp Ala Leu Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe Ser
95 100 105

GAT AGC CCT TTG CTG AAT AGC CTG GTG GGA CAC ATC TTG CAT TCC TCA 1226
Asp Ser Pro Leu Leu Asn Ser Leu Val Gly His Ile Leu His Ser Ser
110 115 120

ATT CTT GTG CCA TAC CAT GGA TGG AGA ATT AGC CAC AGA ACT CAC CAT 1274
Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His Arg Thr His His
125 130 135 140

CAA AAC CAT GGA CAC ATT GAG AAG GAT GAG TCA TGG GTT CCA TTA ACA 1322
Gln Asn His Gly His Ile Glu Lys Asp Glu Ser Trp Val Pro Leu Thr
145 150 155

GAG AAG ATT TAC AAG AAT CTA GAC AGC ATG ACA AGA CTC ATT AGA TTC 1370
Glu Lys Ile Tyr Lys Asn Leu Asp Ser Met Thr Arg Leu Ile Arg Phe
160 165 170

ACT GTG CCA TTT CCA TTG TTT GTG TAT CCA ATT TAT TTG TTT TCA AGA 1418
Thr Val Pro Phe Pro Leu Phe Val Tyr Pro Ile Tyr Leu Phe Ser Arg
175 180 185

AGC CCC GGA AAG GAA GGC TCT CAC TTC AAT CCC TAC AGC AAT CTG TTC 1466
Ser Pro Gly Lys Glu Gly Ser His Phe Asn Pro Tyr Ser Asn Leu Phe
190 195 200

CCA CCC AGT GAG AGA AAA GGA ATA GCA ATA TCA ACA CTG TGT TGG GCT 1514
Pro Pro Ser Glu Arg Lys Gly Ile Ala Ile Ser Thr Leu Cys Trp Ala
205 210 215 220

ACC ATG TTT TCT CTG CTT ATC TAT CTC TCA TTC ATA ACT AGT CCA CTT 1562
Thr Met Phe Ser Leu Leu Ile Tyr Leu Ser Phe Ile Thr Ser Pro Leu
225 230 235

CTA GTG CTC AAG CTC TAT GGA ATT CCA TAT TGG ATA TTT GTT ATG TGG 1610
Leu Val Leu Lys Leu Tyr Gly Ile Pro Tyr Trp Ile Phe Val Met Trp
240 245 250

CTG GAC TTT GTC ACA TAC TTG CAT CAC CAT GGT CAC CAC CAG AAA CTG 1658
Leu Asp Phe Val Thr Tyr Leu His His His Gly His His Gln Lys Leu
255 260 265

CCT TGG TAC CGC GGC AAG GAA TGG AGT TAT TTA AGA GGT GGC CTC ACC 1706
Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr
270 275 280

ACT GTG GAT CGT GAC TAT GGT TGG ATC TAT AAC ATT CAC CAT GAC ATT 1754
Thr Val Asp Arg Asp Tyr Gly Trp Ile Tyr Asn Ile His His Asp Ile
285 290 295 300

GGC ACC CAT GTT ATC CAC CAT CTT TTC CCC CAA ATT CCT CAT TAT CAC 1802
Gly Thr His Val Ile His His Leu Phe Pro Gln Ile Pro His Tyr His
305 310 315

CTC GTT GAA GCG ACA CAA GCA GCA AAA CCA GTT CTT GGA GAT TAC TAC 1850
Leu Val Glu Ala Thr Gln Ala Ala Lys Pro Val Leu Gly Asp Tyr Tyr
320 325 330

CGT GAG CCA GAA AGA TCT GCG CCA TTA CCA TTT CAT CTA ATA AAG TAT 1898
Arg Glu Pro Glu Arg Ser Ala Pro Leu Pro Phe His Leu Ile Lys Tyr
335 340 345

TTA ATT CAG AGT ATG AGA CAA GAC CAC TTC GTA AGT GAC ACT GGA GAT 1946
Leu Ile Gln Ser Met Arg Gln Asp His Phe Val Ser Asp Thr Gly Asp
350 355 360

GTT GTT TAT TAT CAG ACT GAT TCT CTG CTC CTC CAC TCG CAA CGA GAC 1994
Val Val Tyr Tyr Gln Thr Asp Ser Leu Leu Leu His Ser Gln Arg Asp
365 370 375 380

TGAGTTTCAA ACTTTTTGGG TTATTATTTA TTGGATTCTA GCTACTCAAA TTACTTTTTT 2054

TTTAATGTTA TGTTTTTTGG AGTTTAACGT TTTCTGAACA ACTTGCAAAT TACTTGCATA 2114

GAGAGACATG GAATATTTAT TTGAAATTAG TAAGGTAGTA ATAATAAATT TTGAATTGTC 2174

AGTTTCA 2181

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 380 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Val Lys Asp Thr Lys Pro Leu Ala Tyr Ala Ala Asn Asn Gly Tyr
1 5 10 15

Gln Gln Lys Gly Ser Ser Phe Asp Phe Asp Pro Ser Ala Pro Pro Pro
20 25 30

Phe Lys Ile Ala Glu Ile Arg Ala Ser Ile Pro Lys His Cys Trp Val
35 40 45

Lys Asn Pro Trp Arg Ser Leu Ser Tyr Val Leu Arg Asp Val Leu Val
50 55 60

Ile Ala Ala Leu Val Ala Ala Ala Ile His Phe Asp Asn Trp Leu Leu
65 70 75 80

Trp Leu Ile Tyr Cys Pro Ile Gln Gly Thr Met Phe Trp Ala Leu Phe
85 90 95

Val Leu Gly His Asp Cys Gly His Gly Ser Phe Ser Asp Ser Pro Leu
100 105 110

Leu Asn Ser Leu Val Gly His Ile Leu His Ser Ser Ile Leu Val Pro
115 120 125

Tyr His Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His Gly
130 135 140

His Ile Glu Lys Asp Glu Ser Trp Val Pro Leu Thr Glu Lys Ile Tyr
145 150 155 160

Lys Asn Leu Asp Ser Met Thr Arg Leu Ile Arg Phe Thr Val Pro Phe
165 170 175

Pro Leu Phe Val Tyr Pro Ile Tyr Leu Phe Ser Arg Ser Pro Gly Lys
180 185 190

Glu Gly Ser His Phe Asn Pro Tyr Ser Asn Leu Phe Pro Pro Ser Glu
195 200 205

Arg Lys Gly Ile Ala Ile Ser Thr Leu Cys Trp Ala Thr Met Phe Ser
210 215 220

Leu Leu Ile Tyr Leu Ser Phe Ile Thr Ser Pro Leu Leu Val Leu Lys
225 230 235 240

Leu Tyr Gly Ile Pro Tyr Trp Ile Phe Val Met Trp Leu Asp Phe Val
245 250 255

Thr Tyr Leu His His His Gly His His Gln Lys Leu Pro Trp Tyr Arg
260 265 270

Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Val Asp Arg
275 280 285

Asp Tyr Gly Trp Ile Tyr Asn Ile His His Asp Ile Gly Thr His Val
290 295 300

Ile His His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Glu Ala
305 310 315 320

Thr Gln Ala Ala Lys Pro Val Leu Gly Asp Tyr Tyr Arg Glu Pro Glu
325 330 335

Arg Ser Ala Pro Leu Pro Phe His Leu Ile Lys Tyr Leu Ile Gln Ser
340 345 350

Met Arg Gln Asp His Phe Val Ser Asp Thr Gly Asp Val Val Tyr Tyr
355 360 365

Gln Thr Asp Ser Leu Leu Leu His Ser Gln Arg Asp
370 375 380

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1675 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Glycine max

(vii) IMMEDIATE SOURCE:

(B) CLONE: pSFD-118bwp

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 169..1530

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
CTGTGGCAAT TTTTCTCTTC TCCTTCTGGT TCTCATCTTT GTGTTCTTCT TTGTTTCTCA      60

CCTTTCTGAG GATTTTTCCA TCTTAGTTCC TGGAGGCACC AGGAACCTGA CCAAATAAAT     120

AAACCTTTTT TTTCTTCTAA TTTTTCTGAA GTTTCATTTT TTAGTCCA ATG GCA ACT      177
                                                     Met Ala Thr
                                                       1

TGG TAT CAT CAG AAA TGT GGC TTG AAG CCT CTT GCT CCA GTA ATT CCT       225
Trp Tyr His Gln Lys Cys Gly Leu Lys Pro Leu Ala Pro Val Ile Pro
  5                  10                  15

AGA CCT AGA ACT GGG GCT GCT TTG TCC AGC ACC TCA AGG GTT GAA TTT       273
Arg Pro Arg Thr Gly Ala Ala Leu Ser Ser Thr Ser Arg Val Glu Phe
 20                  25                  30                  35

TTG GAC ACA AAC AAG GTA GTG GCA GGT CCT AAG TTT CAA CCT TTG AGG       321
Leu Asp Thr Asn Lys Val Val Ala Gly Pro Lys Phe Gln Pro Leu Arg
                 40                  45                  50

TGC AAC CTC AGG GAG AGG AAT TGG GGG CTG AAA GTG AGT GCC CCT TTG       369
Cys Asn Leu Arg Glu Arg Asn Trp Gly Leu Lys Val Ser Ala Pro Leu
             55                  60                  65

AGG GTT GCT TCC ATT GAA GAG GAG CAA AAG AGT GTT GAT TTA ACC AAT       417
Arg Val Ala Ser Ile Glu Glu Glu Gln Lys Ser Val Asp Leu Thr Asn
         70                  75                  80

GGG ACT AAT GGG GTT GAG CAT GAG AAG CTT CCA GAA TTT GAC CCT GGT       465
Gly Thr Asn Gly Val Glu His Glu Lys Leu Pro Glu Phe Asp Pro Gly
     85                  90                  95

GCT CCG CCA CCA TTC AAC TTG GCT GAT ATT AGA GCA GCC ATT CCA AAG       513
Ala Pro Pro Pro Phe Asn Leu Ala Asp Ile Arg Ala Ala Ile Pro Lys
100                 105                 110                 115
```

```
CAT TGC TGG GTG AAG GAC CCT TGG AGG TCC ATG AGC TAT GTG GTG AGG      561
His Cys Trp Val Lys Asp Pro Trp Arg Ser Met Ser Tyr Val Val Arg
                120                 125                 130

GAT GTG ATT GCT GTC TTT GGT TTG GCT GCT GCT GCT GCG TAT CTC AAT      609
Asp Val Ile Ala Val Phe Gly Leu Ala Ala Ala Ala Ala Tyr Leu Asn
            135                 140                 145

AAT TGG TTG GTT TGG CCT CTC TAT TGG GCT GCT CAA GGC ACT ATG TTC      657
Asn Trp Leu Val Trp Pro Leu Tyr Trp Ala Ala Gln Gly Thr Met Phe
        150                 155                 160

TGG GCT CTG TTT GTT CTT GGT CAT GAT TGT GGT CAT GGA AGC TTT TCA      705
Trp Ala Leu Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe Ser
    165                 170                 175

AAC AAC TCC AAA TTG AAC AGT GTT GTT GGA CAT CTG CTG CAT TCT TCA      753
Asn Asn Ser Lys Leu Asn Ser Val Val Gly His Leu Leu His Ser Ser
180                 185                 190                 195

ATT CTA GTG CCA TAT CAT GGA TGG AGA ATC AGT CAT AGG ACT CAT CAC      801
Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His Arg Thr His His
                200                 205                 210

CAA CAT CAT GGT CAT GCT GAA AAT GAT GAA TCA TGG CAT CCG TTG CCT      849
Gln His His Gly His Ala Glu Asn Asp Glu Ser Trp His Pro Leu Pro
            215                 220                 225

GAA AAA TTG TTC AGA AGC TTG GAC ACT GTA ACT CGT ATG TTA AGA TTC      897
Glu Lys Leu Phe Arg Ser Leu Asp Thr Val Thr Arg Met Leu Arg Phe
        230                 235                 240

ACA GCA CCT TTT CCA CTT CTT GCA TTT CCT GTG TAC CTT TTT AGT AGG      945
Thr Ala Pro Phe Pro Leu Leu Ala Phe Pro Val Tyr Leu Phe Ser Arg
    245                 250                 255

AGT CCT GGG AAG ACT GGT TCT CAC TTT GAC CCC AGC AGT GAC TTG TTC      993
Ser Pro Gly Lys Thr Gly Ser His Phe Asp Pro Ser Ser Asp Leu Phe
260                 265                 270                 275

GTT CCC AAT GAA AGA AAA GAT GTT ATT ACT TCC ACA GCT TGT TGG GCT     1041
Val Pro Asn Glu Arg Lys Asp Val Ile Thr Ser Thr Ala Cys Trp Ala
                280                 285                 290

GCT ATG TTG GGA TTG CTT GTT GGA TTG GGG TTT GTA ATG GGT CCA ATT     1089
Ala Met Leu Gly Leu Leu Val Gly Leu Gly Phe Val Met Gly Pro Ile
            295                 300                 305

CAA CTT CTT AAG CTT TAT GGT GTT CCC TAT GTT ATA TTC GTT ATG TGG     1137
Gln Leu Leu Lys Leu Tyr Gly Val Pro Tyr Val Ile Phe Val Met Trp
        310                 315                 320

TTG GAT TTG GTG ACT TAT TTG CAC CAT CAT GGC CAT GAA GAC AAA TTA     1185
Leu Asp Leu Val Thr Tyr Leu His His His Gly His Glu Asp Lys Leu
    325                 330                 335
```

```
CCT TGG TAC CGT GGA AAG GAA TGG AGC TAC CTC AGG GGT GGT CTA ACT      1233
Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr
340                 345                 350                 355

ACT CTT GAT CGT GAT TAT GGA TGG ATC AAT AAC ATT CAC CAT GAC ATT      1281
Thr Leu Asp Arg Asp Tyr Gly Trp Ile Asn Asn Ile His His Asp Ile
                360                 365                 370

GGC ACT CAT GTC ATT CAT CAC CTA TTT CCT CAA ATT CCA CAC TAT CAC      1329
Gly Thr His Val Ile His His Leu Phe Pro Gln Ile Pro His Tyr His
            375                 380                 385

TTA GTT GAG GCT ACT GAG GCT GCT AAG CCA GTG TTT GGA AAA TAT TAT      1377
Leu Val Glu Ala Thr Glu Ala Ala Lys Pro Val Phe Gly Lys Tyr Tyr
        390                 395                 400

AGA GAA CCA AAG AAA TCA GCA GCA CCT CTT CCT TTT CAC CTT ATT GGG      1425
Arg Glu Pro Lys Lys Ser Ala Ala Pro Leu Pro Phe His Leu Ile Gly
    405                 410                 415

GAA ATA ATA AGG AGC TTC AAG ACT GAC CAT TTT GTT AGT GAC ACG GGG      1473
Glu Ile Ile Arg Ser Phe Lys Thr Asp His Phe Val Ser Asp Thr Gly
420                 425                 430                 435

GAT GTT GTG TAC TAT CAA ACC GAC TCT AAG ATT AAT GGC TCT TCC AAA      1521
Asp Val Val Tyr Tyr Gln Thr Asp Ser Lys Ile Asn Gly Ser Ser Lys
                440                 445                 450

TTA GAG TGAATATTAA AATTCTTTTC TATATAGACA AGAGAGGCTT ATACACAATT       1577
Leu Glu

CTTATTGCTT TAAAGATTGT CTTGAGTTTC TCCGAAAGTT ACTGCACTTA CTTGGAGTTG    1637

AATCCTTCAT TAATAAAGGG ATGGATGGAT CATATAAA                            1675
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 453 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
Met Ala Thr Trp Tyr His Gln Lys Cys Gly Leu Lys Pro Leu Ala Pro
 1               5                  10                  15

Val Ile Pro Arg Pro Arg Thr Gly Ala Ala Leu Ser Ser Thr Ser Arg
            20                  25                  30

Val Glu Phe Leu Asp Thr Asn Lys Val Val Ala Gly Pro Lys Phe Gln
        35                  40                  45
```

Pro Leu Arg Cys Asn Leu Arg Glu Arg Asn Trp Gly Leu Lys Val Ser
50 55 60

Ala Pro Leu Arg Val Ala Ser Ile Glu Glu Glu Gln Lys Ser Val Asp
65 70 75 80

Leu Thr Asn Gly Thr Asn Gly Val Glu His Glu Lys Leu Pro Glu Phe
85 90 95

Asp Pro Gly Ala Pro Pro Pro Phe Asn Leu Ala Asp Ile Arg Ala Ala
100 105 110

Ile Pro Lys His Cys Trp Val Lys Asp Pro Trp Arg Ser Met Ser Tyr
115 120 125

Val Val Arg Asp Val Ile Ala Val Phe Gly Leu Ala Ala Ala Ala Ala
130 135 140

Tyr Leu Asn Asn Trp Leu Val Trp Pro Leu Tyr Trp Ala Ala Gln Gly
145 150 155 160

Thr Met Phe Trp Ala Leu Phe Val Leu Gly His Asp Cys Gly His Gly
165 170 175

Ser Phe Ser Asn Asn Ser Lys Leu Asn Ser Val Val Gly His Leu Leu
180 185 190

His Ser Ser Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His Arg
195 200 205

Thr His His Gln His His Gly His Ala Glu Asn Asp Glu Ser Trp His
210 215 220

Pro Leu Pro Glu Lys Leu Phe Arg Ser Leu Asp Thr Val Thr Arg Met
225 230 235 240

Leu Arg Phe Thr Ala Pro Phe Pro Leu Leu Ala Phe Pro Val Tyr Leu
245 250 255

Phe Ser Arg Ser Pro Gly Lys Thr Gly Ser His Phe Asp Pro Ser Ser
260 265 270

Asp Leu Phe Val Pro Asn Glu Arg Lys Asp Val Ile Thr Ser Thr Ala
275 280 285

Cys Trp Ala Ala Met Leu Gly Leu Leu Val Gly Leu Gly Phe Val Met
290 295 300

Gly Pro Ile Gln Leu Leu Lys Leu Tyr Gly Val Pro Tyr Val Ile Phe
305 310 315 320

Val Met Trp Leu Asp Leu Val Thr Tyr Leu His His His Gly His Glu
325 330 335

Asp Lys Leu Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly
340 345 350

Gly Leu Thr Thr Leu Asp Arg Asp Tyr Gly Trp Il Asn Asn Ile His
355 360 365

His Asp Il Gly Thr His Val Ile His His Leu Phe Pro Gln Ile Pr
370 375 380

His Tyr His Leu Val Glu Ala Thr Glu Ala Ala Lys Pro Val Phe Gly
385 390 395 400

Lys Tyr Tyr Arg Glu Pro Lys Lys Ser Ala Ala Pro Leu Pro Phe His
405 410 415

Leu Ile Gly Glu Ile Ile Arg Ser Phe Lys Thr Asp His Phe Val Ser
420 425 430

Asp Thr Gly Asp Val Val Tyr Tyr Gln Thr Asp Ser Lys Ile Asn Gly
435 440 445

Ser Ser Lys Leu Glu
450

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 396 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Zea mays

(vii) IMMEDIATE SOURCE:

(B) CLONE: pPCR20

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 31..363

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GGATCCACGC ATCATCAGAA TCACGGTCAC ATCCACAGGG ACGAGTCATG GCACCCGATC 60

ACGGAGAAGC TGTACCGGCA ACTAGAGCCA CGCACCAAGA AGCTGAGATT CACGGTGCCC 120

TTCCCCCTGC TCGCATTCCC CGTCTACCTC TTGTACAGGA GCCCCGGCAA GCTCGGCTCC 180

CACTTCCTTC CCAGCAGCGA CCTGTTCAGC CCCAAGGAGA AGAGCGACGT CATGGTGTCA 240

ACCACCTGCT GGTGCATCAT GCTCGCCTCC CTCCTCGCCA TGGCGTGCGC GTTCGGCCCA 300

CTCCAGGTGC TCAAGATGTA CGGCATCCCA TACCTGGTGT TCGTGATGTG GCTTGACCTG 360

GTGACGTACT TACATCACCA CGGCCACGAT GGATCC 396

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 126 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Zea mays

(vii) IMMEDIATE SOURCE:

(B) CLONE: pPCR20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

His His Gln Asn His Gly His Ile His Arg Asp Glu Ser Trp His Pro
1 5 10 15

Ile Thr Glu Lys Leu Tyr Arg Gln Leu Glu Pro Arg Thr Lys Lys Leu
20 25 30

Arg Phe Thr Val Pro Phe Pro Leu Leu Ala Phe Pro Val Tyr Leu Leu
35 40 45

Tyr Arg Ser Pro Gly Lys Leu Gly Ser His Phe Leu Pro Ser Ser Asp
50 55 60

Leu Phe Ser Pro Lys Glu Lys Ser Asp Val Met Val Ser Thr Thr Cys
65 70 75 80

Trp Cys Ile Met Leu Ala Ser Leu Leu Ala Met Ala Cys Ala Phe Gly
85 90 95

Pro Leu Gln Val Leu Lys Met Tyr Gly Ile Pro Tyr Leu Val Phe Val
100 105 110

Met Trp Leu Asp Leu Val Thr Tyr Leu His His His Gly His
115 120 125

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 472 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Arabidopsis thaliana

(vii) IMMEDIATE SOURCE:

(B) CLONE: pFadx-2 and pYacp7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCTCGAGCTA CGTCAGGGCT AAAACCAGGA ACTGGGCATT GAATGTGGCA ACACCTTTAA 60

CAACTCTTCA GTCTCCATCC GAGGAAGACA GGGAGAGATT CGACCCAGGT GCGCCTCCTC 120

CCTTCAATTT GGCGGATATA AGAGCAGCCA TACCTAAGCA TTGTTGGGTT AAGAATCCAT 180

GGATGTCTAT GAGTTATGTT GTCAGAGATG TTGCTATCGT CTTTGGATTG GCTGCTGTTG 240

CTGCTTACTT CAACAATTGG CTTCTCTGGC CTCTCTACTG GTTCGCTCAA GGAACCATGT 300

TCTGGGCTCT CTTTGTCCTT GGCCATGACT GCGGACATGG TAGCTTCTCG AATGATCCGA 360

GGCTGAACAG TGTGGCTGGT CATCTTCTTC ATTCCTCAAT CCTGGTCCCT TACCATGGCT 420

GGAGGATTAG CCACAGAACT CACCACCAGA ACCATGGTCA TGTCGAGAAT GA 472

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 156 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: YES

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Arabidopsis thaliana

(vii) IMMEDIATE SOURCE:

(B) CLONE: pFadx-2 and pYacp7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
Ser Ser Tyr Val Arg Ala Lys Thr Arg Asn Trp Ala Leu Asn Val Ala
1               5                   10                  15

Thr Pro Leu Thr Thr Leu Gln Ser Pro Ser Glu Glu Asp Arg Glu Arg
            20                  25                  30

Phe Asp Pro Gly Ala Pro Pro Pro Phe Asn Leu Ala Asp Ile Arg Ala
        35                  40                  45

Ala Ile Pro Lys His Cys Trp Val Lys Asn Pro Trp Met Ser Met Ser
    50                  55                  60

Tyr Val Val Arg Asp Val Ala Ile Val Phe Gly Leu Ala Ala Val Ala
65                  70                  75                  80

Ala Tyr Phe Asn Asn Trp Leu Leu Trp Pro Leu Tyr Trp Phe Ala Gln
                85                  90                  95

Gly Thr Met Phe Trp Ala Leu Phe Val Leu Gly His Asp Cys Gly His
            100                 105                 110

Gly Ser Phe Ser Asn Asp Pro Arg Leu Asn Ser Val Ala Gly His Leu
        115                 120                 125

Leu His Ser Ser Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His
    130                 135                 140

Arg Thr His His Gln Asn His Gly His Val Glu Asn
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: misc feature
(B) LOCATION: 1..11
(D) OTHER INFORMATION: /note= "N= INOSINE"

(ix) FEATURE:

(A) NAME/KEY: misc feature
(B) LOCATION: 12..31
(D) OTHER INFORMATION: /note= "N= A OR T OR G OR C"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CGGGATCCAC NCAYCAYCAR AAYCAYGGNC A 31

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: misc feature
(B) LOCATION: 1..15
(D) OTHER INFORMATION: /note= "N= INOSINE"

(ix) FEATURE:

(A) NAME/KEY: misc feature
(B) LOCATION: 16..35
(D) OTHER INFORMATION: /note= "N= A OR T OR G OR C"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CGGGATCCRT CRTGNCCRTG RTGRTGNARR TANGT 35

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: misc feature
(B) LOCATION: 1..36
(D) OTHER INFORMATION: /note= "N= INOSINE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TTCGTNNTNG GNCAYGAYTG YGGNCAYGGN CAYGGNAGNT TC 42

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: misc feature
(B) LOCATION: 1..36
(D) OTHER INFORMATION: /note= "N= INOSINE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TTCGTNNTNG GNCAYGAYTG YGGNCAYGGN TCNTTC 36

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGHCAYGAYT GYGGHCAC 18

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GGHCAYGAYT GYGGHCAT 18

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GTACTRTARC CDTGDGTR 18

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GTGCTRTARC CDTGDGTR 18

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GTRCANTARG TRGTRAAYAA YGG 23

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GTRCANTADG TRGTRGADAA YGG 23

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: misc feature
(B) LOCATION: 1..36
(D) OTHER INFORMATION: /note= "N= INOSINE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TTCGTNNTNG GNCAYGAYTG YGGNCAYGGN AGNTTT 36

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: misc feature
(B) LOCATION: 1..36
(D) OTHER INFORMATION: /note= "N= INOSINE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TTCGTNNTNG GNCAYGAYTG YGGNCAYGGN TCNTTT 36

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: misc feature
(B) LOCATION: 1..38
(D) OTHER INFORMATION: /note= "N= INOSINE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTRCTRTANC CNTGNGTNCA NTANGTAGTG RANAAGGG 38

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: misc feature
(B) LOCATION: 1..38

(D) OTHER INFORMATION: /note= "N= INOSINE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GTRCTRTANC CNTGNGTNCA NTANGTGGTG RANAAGGG 38

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 138 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: misc feature
(B) LOCATION: 1..135
(D) OTHER INFORMATION: /note= "N= INOSINE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GTGGTGNGTN CNGTGNGANA NNCKCCANCC GTGGTANGGN ACNANNANGA ANGANGAGTG 60

NANNANGTGN CCNACNANNG AGTTNANNAN NGGNATNTCN GAGAANGANC CGTGNCCGCA 120

NTCGTGNCCN ANNACGAA 138

CLAIMS

1. An isolated nucleic acid fragment comprising a nucleic acid sequence encoding a fatty acid desaturase or a fatty acid desaturase-related enzyme with an amino acid identity of 50% or greater to the polypeptide encoded by SEQ ID NOS:1, 4, 6, 8, 10, 12, 14 or 16.

2. The isolated nucleic acid fragment of Claim 1 wherein the amino acid identity is 65% or greater to the polypeptide encoded by SEQ ID NOS:1, 4, 6, 8, 10, 12, 14 or 16.

3. The isolated nucleic acid fragment of Claim 1 wherein the nucleic acid identity is 90% or greater to SEQ ID NOS:1, 4, 6, 8, 10, 12, 14 or 16.

4. An isolated nucleic acid fragment of Claim 1 wherein said fragment is isolated from a plant selected from the group consisting of soybean, oilseed *Brassica* species, *Arabidopsis thaliana* and corn.

5. A chimeric gene capable of causing altered levels of linolenic acid in a transformed plant cell, the gene comprising a nucleic acid fragment of any of Claims 1, 2, or 3, the fragment operably linked to suitable regulatory sequences.

6. Plants containing the chimeric genes of Claim 5.

7. Oil obtained from seeds of the plants containing the chimeric genes of Claim 5.

8. A method of producing seed oil containing altered levels of linolenic (18:3) acid comprising:

(a) transforming a plant cell of an oil-producing species with a chimeric gene of Claim 5;

(b) growing fertile plants from the transformed plant cells of step (a);

(c) screening progeny seeds from the fertile plants of step (b) for the desired levels of linolenic (18:3) acid; and

(d) processing the progeny seed of step (c) to obtain seed oil containing altered levels of linolenic (18:3) acid.

9 The product of the method of Claim 8.

10. A method of Claim 8 wherein said plant cell of an oil-producing species is selected from the group consisting of Arabidopsis thaliana, soybean, oilseed Brassica species, sunflower, cotton, cocoa, peanut, safflower, and corn.

11. A method of breeding plant species producing altered levels of linolenic acid in the seed oil of oil-producing plant species comprising:

(a) making a cross between two varieties of oil-producing species differing in the linolenic acid trait;

(b) making a Southern blot of restriction enzyme digested genomic DNA isolated from several progeny plants resulting from the cross of step (a); and

(c) hybridizing the Southern blot with a radiolabelled nucleic acid fragment of Claim 1.

12. The product of the method of Claim 11.

13. A method of RFLP mapping in a genomic RFLP marker comprising:

(a) making a cross between two varieties of plants;

(b) making a Southern blot of restriction enzyme digested genomic DNA isolated from several progeny plants resulting from the cross of step (a); and

(c) hybridizing the Southern blot with a radiolabelled nucleic acid fragments of Claim 1.

14. A method to isolate nucleic acid fragments encoding fatty acid desaturases and fatty acid desaturase-related enzymes, comprising:

(a) comparing SEQ ID NOS:2, 5, 7, 9, 11, 13, 15 and 17 with other fatty acid desaturase polypeptide sequences;

(b) identifying the conserved sequence(s) of 4 or more amino acids obtained in step (a);

(c) making region-specific nucleotide probe(s) or oligomer(s) based on the conserved sequences identified in step b; and

(d) using the nucleotide probe(s) or oligomers(s) of step c to isolate sequences encoding fatty acid desaturases and fatty-acid desaturase-related enzymes by sequence-dependent protocols.

15. The product of the method of Claim 14.

16. The isolated genomic DNA of Arabidopsis thaliana identified by accession number ATCC 75167.

17. An isolated cDNA clone which encodes for soybean delta-15 desaturase, the clone designated pXF1 comprising the DNA sequence of SEQ ID NO 10 and identified by accession number ATCC 68874.

18. An isolated cDNA clone which encodes for oilseed Brassica species delta-15 desaturase, the clone designated pBNSF3 comprising the DNA sequence of SEQ ID NO:6 and identified by accession number ATCC 68854.

19. An isolated Polymerase Chain Reaction Product for Zea mays delta-15 desaturase, the clone designated pcr20 comprising the DNA sequence of SEQ ID NO:14.

INTERNATIONAL SEARCH REPORT

International Application No ?CT/US 92/10284

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]
According to International Patent Classification (IPC) or to both National Classification and IPC
Int.Cl. 5 C12N15/53; C12N15/82; C11B1/00; C12Q1/68

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12N ; C11B ; C12Q

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category [o]	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
X	UCLA SYMP. MOL. CELL. BIOL.; NEW SER., PLANT GENE TRANSFER vol. 129, 1990, pages 301 - 309 BROWSE, J., ET AL. 'Strategies for modifying plant lipid composition'	7,11,14
Y	see the whole document	2-6,8, 10,15

Y	SCIENCE vol. 252, 5 April 1991, LANCASTER, PA US pages 80 - 87 SOMERVILLE, C., ET AL. 'Plant lipids: Metabolism, mutants, and membranes' see page 82, right column, line 24 - line 27	2-6,8, 10,15

	-/--	

° Special categories of cited documents : [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
17 MARCH 1993	29. 03. 93
International Searching Authority	**Signature of Authorized Officer**
EUR PEAN PATENT FFICE	MADDOX A.D.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
X	THEOR. APPL. GENET. vol. 80, no. 2, 1990, pages 234 - 240 LEMIEUX, B., ET AL. 'Mutants of Arabidopsis with alterations in seed lipid fatty acid composition' see the whole document	7,9,11
P,X	SCIENCE vol. 258, 20 November 1992, LANCASTER, PA US pages 1353 - 1355 ARONDEL, V., ET AL. 'Map-based cloning of a gene controlling omega-3 fatty acid desaturation in Arabidopsis' see the whole document	1-13,15
A	PLANT PHYSIOLOGY. vol. 81, no. 3, 1986, ROCKVILLE, MD, USA. pages 859 - 864 BROWSE, J., ET AL. 'A mutant of Arabidopsis deficient in C18:3 and C16:3 leaf lipids' see the whole document	1-12
A	WO,A,9 113 972 (CALGENE) 19 September 1991 see the whole document	1-10

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO.

US 9210284
SA 67975

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on 17/03/93
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9113972	19-09-91	EP-A- 0472722	04-03-92

EPO FORM P0479

For more details about this annex : see Official Journal of the European Patent ffice, No. 12/82

3

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/53, 15/82, C11B 1/00 C12Q 1/68, A01H 5/00

A1

(11) International Publication Number: WO 94/11516

(43) International Publication Date: 26 May 1994 (26.05.94)

(21) International Application Number: PCT/US93/09987

(22) International Filing Date: 15 October 1993 (15.10.93)

(30) Priority data:
07/977,339 17 November 1992 (17.11.92) US

(60) Parent Application or Grant
(63) Related by Continuation
US 07/977,339 (CIP)
Filed on 17 November 1992 (17.11.92)

(71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only) : LIGHTNER, Jonathan, Edward [US/US]; 438 East Market Street, Marietta, PA 17547 (US). OKULEY, John, Joseph [US/US]; 217 Fallis Road, Columbus, OH 43214 (US).

(74) Agents: MORRISSEY, Bruce, W. et al.; E.I. du Pont de Nemours and Company, Legal/Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States: AU, BR, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: GENES FOR MICROSOMAL DELTA-12 FATTY ACID DESATURASES AND RELATED ENZYMES FROM PLANTS

(57) Abstract

The preparation and use of nucleic acid fragments encoding fatty acid desaturase enzymes are described. The invention permits alteration of plant lipid composition. Chimeric genes incorporating such nucleic acid fragments with suitable regulatory sequences may be used to create transgenic plants with altered levels of unsaturated fatty acids.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SI	Slovenia
SK	Slovakia
SN	Senegal
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

TITLE

GENES FOR MICROSOMAL DELTA-12 FATTY ACID DESATURASES AND RELATED ENZYMES FROM PLANTS

FIELD OF THE INVENTION

The invention relates to the preparation and use of nucleic acid fragments encoding fatty acid desaturase enzymes to modify plant lipid composition. Chimeric genes incorporating such nucleic acid fragments and suitable regulatory sequences may be used to create transgenic plants with altered levels of unsaturated fatty acids.

BACKGROUND OF THE INVENTION

Plant lipids have a variety of industrial and nutritional uses and are central to plant membrane function and climatic adaptation. These lipids represent a vast array of chemical structures, and these structures determine the physiological and industrial properties of the lipid. Many of these structures result either directly or indirectly from metabolic processes that alter the degree of unsaturation of the lipid. Different metabolic regimes in different plants produce these altered lipids, and either domestication of exotic plant species or modification of agronomically adapted species is usually required to economically produce large amounts of the desired lipid.

Plant lipids find their major use as edible oils in the form of triacylglycerols. The specific performance and health attributes of edible oils are determined largely by their fatty acid composition. Most vegetable oils derived from commercial plant varieties are composed primarily of palmitic (16:0), stearic (18:0), oleic (18:1), linoleic (18:2) and linolenic (18:3) acids. Palmitic and stearic acids are, respectively, 16- and 18-carbon-long, saturated fatty acids. Oleic, linoleic, and linolenic acids are 18-carbon-long,

unsaturated fatty acids containing one, two, and three double bonds, respectively. Oleic acid is referred to as a mono-unsaturated fatty acid, while linoleic and linolenic acids are referred to as poly-unsaturated fatty acids. The relative amounts of saturated and unsaturated fatty acids in commonly used, edible vegetable oils are summarized below (Table 1):

TABLE 1

Percentages of Saturated and Unsaturated Fatty Acids in the Oils of Selected Oil Crops

	Saturated	Mono-unsaturated	Poly-unsaturated
Canola	6%	58%	36%
Soybean	15%	24%	61%
Corn	13%	25%	62%
Peanut	18%	48%	34%
Safflower	9%	13%	78%
Sunflower	9%	41%	51%
Cotton	30%	19%	51%

Many recent research efforts have examined the role that saturated and unsaturated fatty acids play in reducing the risk of coronary heart disease. In the past, it was believed that mono-unsaturates, in contrast to saturates and poly-unsaturates, had no effect on serum cholesterol and coronary heart disease risk. Several recent human clinical studies suggest that diets high in mono-unsaturated fat and low in saturated fat may reduce the "bad" (low-density lipoprotein) cholesterol while maintaining the "good" (high-density lipoprotein) cholesterol (Mattson et al., Journal of Lipid Research (1985) 26:194-202).

A vegetable oil low in total saturates and high in mono-unsaturates would provide significant health

benefits to consumers as well as economic benefits to oil processors. As an example, canola oil is considered a very healthy oil. However, in use, the high level of poly-unsaturated fatty acids in canola oil renders the oil unstable, easily oxidized, and susceptible to development of disagreeable odors and flavors (Gailliard, 1980, Vol. 4, pp. 85-116 In: Stumpf, P. K., Ed., The Biochemistry of Plants, Academic Press, New York). The levels of poly-unsaturates may be reduced by hydrogenation, but the expense of this process and the concomitant production of nutritionally questionable trans isomers of the remaining unsaturated fatty acids reduces the overall desirability of the hydrogenated oil (Mensink et al., New England J. Medicine (1990) N323: 439-445). Similar problems exist with soybean and corn oils.

For specialized uses, high levels of poly-unsaturates can be desirable. Linoleate and linolenate are essential fatty acids in human diets, and an edible oil high in these fatty acids can be used for nutritional supplements, for example in baby foods.

Mutation-breeding programs have met with some success in altering the levels of poly-unsaturated fatty acid levels found in the edible oils of agronomic species. Examples of commercially grown varieties are high (85%) oleic sunflower and low (2%) linolenic flax (Knowles, (1980) pp. 35-38 In: Applewhite, T. H., Ed., World Conference on Biotechnology for the Fats and Oils Industry Proceedings, American Oil Chemists' Society). Similar commercial progress with the other plants shown in Table 1 has been largely elusive due to the difficult nature of the procedure and the pleiotropic effects of the mutational regime on plant hardiness and yield potential.

The biosynthesis of the major plant lipids has been the focus of much research (Browse et al., Ann. Rev. Plant Physiol. Mol. Biol. (1991) 42:467-506). These studies show that, with the notable exception of the soluble stearoyl-acyl carrier protein desaturase, the controlling steps in the production of unsaturated fatty acids are largely catalyzed by membrane-associated fatty acid desaturases. Desaturation reactions occur in plastids and in the endoplasmic reticulum using a variety of substrates including galactolipids, sulfolipids, and phospholipids. Genetic and physiological analyses of Arabidopsis thaliana nuclear mutants defective in various fatty acid desaturation reactions indicates that most of these reactions are catalyzed by enzymes encoded at single genetic loci in the plant. The analyses show further that the different defects in fatty acid desaturation can have profound and different effects on the ultra-structural morphology, cold sensitivity, and photosynthetic capacity of the plants (Ohlrogge, et al., Biochim. Biophys. Acta (1991) 1082:1-26). However, biochemical characterization of the desaturase reactions has been meager. The instability of the enzymes and the intractability of their proper assay has largely limited researchers to investigations of enzyme activities in crude membrane preparations. These investigations have, however, demonstrated the role of delta-12 desaturase and delta-15 desaturase activities in the production of linoleate and linolenate from 2-oleoyl-phosphatidylcholine and 2-linoleoyl-phosphatidylcholine, respectively (Wang et al., Plant Physiol. Biochem. (1988) 26:777-792). Thus, modification of the activities of these enzymes represents an attractive target for altering the levels of lipid unsaturation by genetic engineering.

Nucleotide sequences encoding microsomal delta-9 stearoyl-coenzyme-A desaturases from yeast, rat, and mice have been described (Stukey, et al., J. Biol. Chem.(1990) 265:20144-20149; Thiede, et al., J. Biol. Chem. (1986) 261:13230-13235; Kaestner, et al., J. Biol. Chem. (1989) 264:14755-1476). Nucleotide sequences encoding soluble delta-9 stearoyl-acyl carrier protein desaturases from higher plants have also been described (Thompson, et al., Proc. Natl. Acad. Sci. U.S.A. (1991) 88:2578-2582; Shanklin et al., Proc. Natl. Acad. Sci. USA (1991) 88:2510-2514). A nucleotide sequence from coriander plant encoding a soluble fatty acid desaturase, whose deduced amino acid sequence is highly identical to that of the stearoyl-acyl carrier protein desaturase and which is responsible for introducing the double bond in petroselinic fatty acid (18:1, 6c), has also been described [Cahoon, et. al. (1992) Proc. Natl. Acad. Sci. U.S.A. 89:11184-11188]. Two fatty acid desaturase genes from the cyanobacterium, Synechocystis PCC6803, have been described: one encodes a fatty acid desaturase, designated *des A*, that catalyzes the conversion of oleic acid at the sn-1 position of galactolipids to linoleic acid [Wada, et al., Nature (1990) 347:200-203]; another encodes a delta-6 fatty acid desaturase that catalyzes the conversion of linoleic acid at the sn-1 position of galactolipids to γ-linolenic acid (18:2, 6c,9c) [WO 9306712]. Nucleotide sequences encoding higher plant membrane-bound microsomal and plastid delta-15 fatty acid desaturases have also been described [WO 9311245]; Arondel, V. et. al. (1992) Science 258:1353-1355]. There is no report of the isolation of higher plant genes encoding fatty acid desaturases other than the soluble delta-6 and delta-9 desaturases and the membrane-bound (microsomal and plastid) delta-15 desaturases. While there is

extensive amino acid sequence identity between the soluble desaturases and significant amino acid sequence identity between the higher plant microsomal and plastid delta-15 desaturases, there is no significant homology between the soluble and the membrane-bound desaturases. Sequence-dependent protocols based on the sequences encoding delta-15 desaturases have been unsuccessful in cloning sequences for microsomal delta-12 desaturase. For example, nucleotide sequences of microsomal or plastid delta-15 desaturases as hybridization probes have been unsuccessful in isolating a plant microsomal delta-12 desaturase clone. Furthermore, while we have used a set of degenerate oligomers made to a stretch of 12 amino acids, which is identical in all plant delta-15 desaturases and highly conserved (10/12) in the cyanobacterial *des A* desaturase, as a hybridization probe to isolate a higher plant nucleotide sequence encoding plastid delta-12 fatty acid desaturase, this method has been unsuccessful in isolating the microsomal delta-12 desaturase cDNAs. Furthermore, there has been no success in isolating the microsomal delta-12 desaturase by using the polymerase chain reaction products derived from plant DNA, plant RNA or plant cDNA library using PCR primers made to stretches of amino acids that are conserved between the higher plant delta-15 and *des A* desaturases. Thus, there are no teachings which enable the isolation of plant microsomal delta-12 fatty acid desaturases or plant fatty acid desaturase-related enzymes. Furthermore, there is no evidence for a method to control the the level of delta-12 fatty acid desaturation or hydroxlylation in plants using nucleic acids encoding delta-12 fatty acid desaturases or hydroxylases.

The biosynthesis of the minor plant lipids has been less well studied. While hundreds of different fatty

acids have been found, many from the plant kingdom, only a tiny fraction of all plants have been surveyed for their lipid content (Gunstone, et al., Eds., (1986) The Lipids Handbook, Chapman and Hall Ltd., Cambridge). Accordingly, little is known about the biosynthesis of these unusual fatty acids and fatty acid derivatives. Interesting chemical features found in such fatty acids include, for example, allenic and conjugated double bonds, acetylenic bonds, trans double bonds, multiple double bonds, and single double bonds in a wide number of positions and configurations along the fatty acid chain. Similarly, many of the structural modifications found in unusual lipids (e.g., hydroxylation, epoxidation, cyclization, etc.) are probably produced via further metabolism following chemical activation of the fatty acid by desaturation or they involve a chemical reaction that is mechanistically similar to desaturation. Many of these fatty acids and derivatives having such features within their structure could prove commercially useful if an agronomically viable species could be induced to synthesize them by introduction of a gene encoding the appropriate desaturase. Of particular interest are vegetable oils rich in 12-hydroxyoctadeca-9-enoic acid (ricinoleic acid). Ricinoleic acid and its derivatives are widely used in the manufacture of lubricants, polymers, cosmetics, coatings and pharmaceuticals (e.g., see Gunstone, et al., Eds., (1986) The Lipids Handbook, Chapman and Hall Ltd., Cambridge). The only commercial source of ricinoleic acid is castor oil and 100% of the castor oil used by the U.S. is derived from beans grown elsewhere in the world, mainly Brazil. Ricinoleic acid in castor beans is synthesized by the addition of an hydroxyl group at the delta-12 position of oleic acid (Galliard & Stumpf (1966) J. Biol. Chem. 241: 5806-5812). This reaction

resembles the initial reaction in a possible mechanism for the desaturation of oleate at the delta-12 position to linoleate since dehydration of 12-hydroxyoctadeca-9-enoic acid, by an enzyme activity analogous to the hydroxydecanoyl dehydrase of E. coli (Cronan et al. (1988) J. Biol. Chem. 263:4641-4646), would result in the formation of linoleic acid. Evidence for the hydroxylation reaction being part of a general mechanism of enzyme-catalyzed desaturation in eukaryotes has been obtained by substituting a sulfur atom in the place of carbon at the delta-9 position of stearic acid. When incubated with yeast cell extracts the thiostearate was converted to a 9-sulfoxide (Buist et al. (1987) Tetrahedron Letters 28:857-860). This sulfoxidation was specific for sulfur at the delta-9 position and did not occur in a yeast delta-9-desaturase deficient mutant (Buist & Marecak (1991) Tetrahedron Letters 32:891-894). The 9-sulfoxide is the sulfur analogue of 9-hydroxyoctadecastearate, the proposed intermediate of stearate desaturation.

Hydroxylation of oleic acid to ricinoleic acid in castor bean cells, like microsomal desaturation of oleate in plants, occurs at the delta-12 position of the fatty acid at the sn-2 position of phosphatidylcholine in microsomes (Bafor et al. (1991) Plant Physiol 280:507-514). Furthermore, castor oleate delta-12 hydroxylation and plant oleate microsomal delta-12 desaturation are both inhibited by iron chelators and require molecular oxygen [Moreau & Stumpf (1981) Plant Physiology 67:672-676; Somerville, C. (1992) MSU-DOE Plant Research Laboratory Annual Report]. These biochemical similarities in conjunction with the observation that antibodies raised against cytochrome b_5 completely inhibit the activities of both oleate delta-12 desaturation in safflower microsomes and oleate

delta-12 hydroxylase in castor microsomes [Somerville, C. (1992) MSU-DOE Plant Research Laboratory Annual Report] comprise strong evidence that the hydroxylase and the desaturase are functionally related. It seems reasonable to assume, therefore, that the nucleotide sequence encoding a plant delta-12 desaturase would be useful in cloning the oleate hydroxylase gene from castor by sequence-dependent protocols. For example, by screening a castor DNA library with oligomers based on amino acid regions conserved between delta-12 desaturases, or regions conserved between delta-12 and other desaturases, or with oligomers based on amino acids conserved between delta-12 desaturases and known membrane-associated hydroxylases. It would be more efficient to isolate the castor oleate hydroxylase cDNA by combining the sequence dependent protocols with a "differential" library approach. One example of such a difference library would be based on different stages of castor seed development, since ricinoleic acid is not synthesized by very young castor seeds (less than 12 DAP, corresponding to stage I and stage II seeds in the scheme of Greenwood & Bewley, Can. J. Bot. (1982) 60:1751-1760), in the 20 days following these early stages the relative ricinoleate content increases from 0% to almost 90% of total seed fatty acids (James et al. Biochem. J. (1965) 95:448-452, Canvin. Can. J. Biochem. Physiol. (1963) 41:1879-1885). Thus it would be possible to make a cDNA "difference" library made from mRNA present in a stage when ricinoleic acid was being synthesized at a high rate but from which mRNA present in earlier stages was removed. For the earlier stage mRNA, a stage such as stage II (10 DAP) when ricinoleic acid is not being made but when other unsaturated fatty acids are, would be appropriate. The construction of libraries containing only differentially expressed genes

is well known in the art (Sargent. Meth. Enzymol. (1987) 152:423-432). Assembly of the free ricinoleic acid, via ricinoleoyl-CoA, into triacylglycerol is readily catalyzed by canola and safflower seed microsomes (Bafor et al., Biochem J. (1991) 280:507-514, Wiberg et al. 10th International Symposium on the Metabolism, Strucure & Function of Plant Lipids (1992), Jerba, Tunisia) and ricinoleic acid is removed from phosphatidylcholine by a lipase common to all oilseeds investigated. Thus, expression of the castor bean oleate hydroxylase gene in oil crops, such as canola seeds and soybeans, would be expected to result in an oil rich in triglycerides containing ricinoleic acid.

SUMMARY OF THE INVENTION

Applicants have discovered a means to control the nature and levels of unsaturated fatty acids in plants. Nucleic acid fragments from cDNAs or genes encoding fatty acid desaturases are used to create chimeric genes. The chimeric genes may be used to transform various plants to modify the fatty acid composition of the plant or the oil produced by the plant. More specifically, one embodiment of the invention is an isolated nucleic acid fragment comprising a nucleotide sequence encoding a fatty acid desaturase or a fatty acid desaturase-related enzyme with an amino acid identity of 50%, 60%, 90% or greater respectively to the polypeptide encoded by SEQ ID NOS:1, 3, 5, 7, 9, 11, or 15. Most specifically, the invention pertains to a gene sequence for plant microsomal delta-12 fatty acid desaturase or desaturase-related enzyme. The plant in this embodiment may more specifically be soybean, oilseed *Brassica* species, *Arabidopsis thaliana*, castor, and corn.

Another embodiment of this invention involves the use of these nucleic acid fragments in sequence-

dependent protocols. Examples include use of the fragments as hybridization probes to isolate nucleotide sequences encoding other fatty acid desaturases or fatty acid desaturase-related enzymes. A related embodiment involves using the disclosed sequences for amplification of RNA or DNA fragments encoding other fatty acid desaturases or fatty acid desaturase-related enzymes.

Another aspect of this invention involves chimeric genes capable of modifying the fatty acid composition in the seed of a transformed plant, the gene comprising nucleic acid fragments related as defined to SEQ ID NOS:1, 3, 5, 7, 9, or 15 encoding fatty acid desaturases or SEQ ID NOS:11 encoding a desaturase or desaturase-related enzyme operably-linked in suitable orientation to suitable regulatory sequences. Preferred are those chimeric genes which incorporate nucleic acid fragments encoding microsomal delta-12 fatty acid desaturase or desaturase-related enzymes.

Yet another embodiment of the invention involves a method of producing seed oil containing altered levels of unsaturated fatty acids comprising: (a) transforming a plant cell with a chimeric gene described above; (b) growing sexually mature plants from the transformed plant cells of step (a); (c) screening progeny seeds from the sexually mature plants of step (b) for the desired levels of unsaturated fatty acids, and (d) processing the progeny seed of step (c) to obtain seed oil containing altered levels of the unsaturated fatty acids. Preferred plant cells and oils are derived from soybean, rapeseed, sunflower, cotton, cocoa, peanut, safflower, coconut, flax, oil palm, and corn. Preferred methods of transforming such plant cells would include the use of Ti and Ri plasmids of Agrobacterium, electroporation, and high-velocity ballistic bombardment.

The invention also is embodied in a method of RFLP bre ding to obtain altered levels of oleic acids in the seed oil of oil producing plant species. This method involves (a) making a cross between two varieties of oil producing plant species differing in the oleic acid trait; (b) making a Southern blot of restriction enzyme digested genomic DNA isolated from several progeny plants resulting from the cross; and (c) hybridizing the Southern blot with the radiolabelled nucleic acid fragments encoding the fatty acid desaturases or desaturase-related enzymes.

The invention is also embodied in a method of RFLP mapping that uses the isolated microsomal delta-12 desaturase cDNA or related genomic fragments described herein.

The invention is also embodied in plants capable of producing altered levels of fatty acid desaturase by virtue of containing the chimeric genes described herein. Further, the invention is embodied by seed oil obtained from such plants.

BRIEF DESCRIPTION OF THE SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the Sequence Descriptions which form a part of this application. The Sequence Descriptions contain the three letter codes for amino acids as defined in 37 C.F.R. 1.822 which are incorporated herein by reference.

SEQ ID NO:1 shows the 5' to 3' nucleotide sequence of 1372 base pairs of the *Arabidopsis thaliana* cDNA which encodes microsomal delta-12 desaturase. Nucleotides 93-95 and nucleotides 1242-1244 are, respectively, the putative initiation codon and the termination codon of the open reading frame (nucleotides 93-1244). Nucleotides 1-92 and 1245-1372 are, respectively, the 5' and 3' untranslated nucleotides.

SEQ ID NO:2 is the 383 amino acid protein sequence deduced from the open reading frame (nucleotides 93-1244 in SEQ ID NO:1.

SEQ ID NO:3 shows the 5' to 3' nucleotide sequence of 1394 base pairs of the Brassica napus cDNA which encodes microsomal delta-12 desaturase in plasmid pCF2-165d. Nucleotides 99 to 101 and nucleotides 1248 to 1250 are, respectively, the putative initiation codon and the termination codon of the open reading frame (nucleotides 99 to 1250). Nucleotides 1 to 98 and 1251 to 1394 are, respectively, the 5' and 3' untranslated nucleotides.

SEQ ID NO:4 is the 383 amino acid protein sequence deduced from the open reading frame (nucleotides 99 to 1250) in SEQ ID NO:3.

SEQ ID NO:5 shows the 5' to 3' nucleotide sequence of 1369 base pairs of soybean (Glycine max) cDNA which encodes microsomal delta-12 desaturase in plasmid pSF2-169K. Nucleotides 108 to 110 and nucleotides 1245 to 1247 are, respectively, the putative initiation codon and the termination codon of the open reading frame (nucleotides 108 to 1247). Nucleotides 1 to 107 and 1248 to 1369 are, respectively, the 5' and 3' untranslated nucleotides.

SEQ ID NO:6 is the 381 amino acid protein sequence deduced from the open reading frame (nucleotides 113 to 1258) in SEQ ID NO:5.

SEQ ID NO:7 shows the 5' to 3' nucleotide sequence of 1790 base pairs of corn (Zea mays) cDNA which encodes microsomal delta-12 desaturase in plasmid pFad2#1. Nucleotides 165 to 167 and nucleotides 1326 to 1328 are, respectively, the putative initiation codon and the termination codon of the open reading frame (nucleotides 164 to 1328). Nucleotides 1 to 163 and 1329 to 1790

are, respectively, the 5' and 3' untranslated nucleotides.

SEQ ID NO:8 is the 387 amino acid protein sequence deduced from the open reading frame (nucleotides 164 to 1328) in SEQ ID NO:7.

SEQ ID NO:9 shows the 5' to 3' nucleotide sequence of 673 base pairs of castor (Ricinus communis) incomplete cDNA which encodes part of a microsomal delta-12 desaturase in plasmid pRF2-1C. The sequence encodes an open reading frame from base 1 to base 673.

SEQ ID NO:10 is the 219 amino acid protein sequence deduced from the open reading frame (nucleotides 1 to 657) in SEQ ID NO:9.

SEQ ID NO:11 shows the 5' to 3' nucleotide sequence of 1369 base pairs of castor (Ricinus communis) cDNA which encodes part of a microsomal delta-12 desaturase or desaturase-related enzyme in plasmid pRF197C-42. Nucleotides 184 to 186 and nucleotides 1340 to 1342 are, respectively, the putative initiation codon and the termination codon of the open reading frame (nucleotides 184 to 1347). Nucleotides 1 to 183 and 1348 to 1369 are, respectively, the 5' and 3' untranslated nucleotides.

SEQ ID NO:12 is the 387 amino acid protein sequence deduced from the open reading frame (nucleotides 184 to 1342) in SEQ ID NO:11.

SEQ ID NO:13 is the sequence of a set of 64-fold degenerate 26 nucleotide-long oligomers, designated NS3, made to conserved amino acids 101-109 of SEQ ID NO:2, designed to be used as sense primers in PCR to isolate novel sequences encoding microsomal delta-12 desaturases or desaturase-like enzymes.

SEQ ID NO:14 is the sequence of a set of 64-fold degenerate and 26 nucleotide-long oligomers, designated NS9, which is made to conserved amino acids 313-321 of

SEQ ID NO:2 and designed to be used as antisense primers in PCR to isolate novel sequences encoding microsomal delta-12 desaturases or desaturase-like enzymes.

SEQ ID NO:15 shows the 5' to 3' nucleotide sequence of 2973 bp of Arabidopsis thaliana genomic fragment containing the microsomal delta-12 desaturase gene contained in plasmid pAGF2-6. Its nucleotides 433 and 2938 correspond to the start and end, respectively, of SEQ ID NO:1. Its nucleotides 521 to 1654 are the 1134 bp intron.

SEQ ID NO:16 is the sequence of a set of 256-fold degenerate and 25 nucleotide-long oligomers, designated RB5a, which is made to conserved amino acids 318-326 of SEQ ID NO:2 and designed to be used as antisense primers in PCR to isolate novel sequences encoding microsomal delta-12 desaturases or desaturase-like enzymes.

SEQ ID NO:17 is the sequence of a set of 128-fold degenerate and 25 nucleotide-long oligomers, designated RB5b, which is made to conserved amino acids 318-326 of SEQ ID NO:2 and designed to be used as antisense primers in PCR to isolate novel sequences encoding microsomal delta-12 desaturases or desaturase-like enzymes.

DETAILED DESCRIPTION OF THE INVENTION

Applicants have isolated nucleic acid fragments that encode plant fatty acid desaturases and that are useful in modifying fatty acid composition in oil-producing species by genetic transformation.

Thus, transfer of the nucleic acid fragments of the invention or a part thereof that encodes a functional enzyme, along with suitable regulatory sequences that direct the transcription of their mRNA, into a living cell will result in the production or over-production of plant fatty acid desaturases and will result in increased levels of unsaturated fatty acids in cellular lipids, including triacylglycerols.

Transfer of the nucleic acid fragments of the invention or a part thereof, along with suitable regulatory sequences that direct the transcription of their antisense RNA, into plants will result in the inhibition of expression of the endogenous fatty acid desaturase that is substantially homologous with the transferred nucleic acid fragment and will result in decreased levels of unsaturated fatty acids in cellular lipids, including triacylglycerols.

Transfer of the nucleic acid fragments of the invention or a part thereof, along with suitable regulatory sequences that direct the transcription of their mRNA, into plants may result in inhibition by cosuppression of the expression of the endogenous fatty acid desaturase gene that is substantially homologous with the transferred nucleic acid fragment and may result in decreased levels of unsaturated fatty acids in cellular lipids, including triacylglycerols.

The nucleic acid fragments of the invention can also be used as restriction fragment length polymorphism (RFLP) markers in plant genetic mapping and plant breeding programs.

The nucleic acid fragments of the invention or oligomers derived therefrom can also be used to isolate other related fatty acid desaturase genes using DNA, RNA, or a library of cloned nucleotide sequences from the same or different species by well known sequence-dependent protocols, including, for example, methods of nucleic acid hybridization and amplification by the polymerase chain reaction.

Definitions

In the context of this disclosure, a number of terms shall be used. Fatty acids are specified by the number of carbon atoms and the number and position of the double bond: the numbers before and after the colon

refer to the chain l ngth and the number of double bonds, respectively. The number following the fatty acid designation indicates the position of the double bond from the carboxyl end of the fatty acid with the "c" affix for the *cis*-configuration of the double bond. For example, palmitic acid (16:0), stearic acid (18:0), oleic acid (18:1,9c), petroselinic acid (18:1, 6c), linoleic acid (18:2,9c,12c), γ-linolenic acid (18:3, 6c,9c,12c) and α-linolenic acid (18:3, 9c,12c,15c). Unless otherwise specified 18:1, 18:2 and 18:3 refer to oleic, linoleic and linolenic fatty acids. Ricinoleic acid refers to an 18 carbon fatty acid with a *cis*-9 double bond and a 12-hydroxyl group. The term "fatty acid desaturase" used herein refers to an enzyme which catalyzes the breakage of a carbon-hydrogen bond and the introduction of a carbon-carbon double bond into a fatty acid molecule. The fatty acid may be free or esterified to another molecule including, but not limited to, acyl-carrier protein, coenzyme A, sterols and the glycerol moiety of glycerolipids. The term "glycerolipid desaturases" used herein refers to a subset of the fatty acid desaturases that act on fatty acyl moieties esterified to a glycerol backbone. "Delta-12 desaturase" refers to a fatty acid desaturase that catalyzes the formation of a double bond between carbon positions 6 and 7 (numbered from the methyl end), (i.e., those that correspond to carbon positions 12 and 13 (numbered from the carbonyl carbon) of an 18 carbon-long fatty acyl chain. "Delta-15 desaturase" refers to a fatty acid desaturase that catalyzes the formation of a double bond between carbon positions 3 and 4 (numbered from the methyl end), (i.e., those that correspond to carbon positions 15 and 16 (numbered from the carbonyl carbon) of an 18 carbon-long fatty acyl chain. Examples of fatty acid desaturases include, but are not limited

to, the microsomal delta-12 and delta-15 desaturases that act on phosphatidylcholine lipid substrat s; the chloroplastic or plastid delta-12 and delta-15 desaturases that act on phosphatidyl glycerol and galactolipids; and other desaturases that act on such fatty acid substrates such as phospholipids, galacto-lipids, and sulfolipids. "Microsomal desaturase" refers to the cytoplasmic location of the enzyme, while "chloroplast desaturase" and "plastid desaturase" refer to the plastid location of the enzyme. These fatty acid desaturases may be found in a variety of organisms including, but not limited to, higher plants, diatoms, and various eukaryotic and prokaryotic microorganisms such as fungi and photosynthetic bacteria and algae. The term "homologous fatty acid desaturases" refers to fatty acid desaturases that catalyze the same desaturation on the same lipid substrate. Thus, microsomal delta-15 desaturases, even from different plant species, are homologous fatty acid desaturases. The term "heterologous fatty acid desaturases" refers to fatty acid desaturases that catalyze desaturations at different positions and/or on different lipid substrates. Thus, for example, microsomal delta-12 and delta-15 desaturases, which act on phosphatidylcholine lipids, are heterologous fatty acid desaturases, even when from the same plant. Similarly, microsomal delta-15 desaturase, which acts on phosphatidylcholine lipids, and chloroplast delta-15 desaturase, which acts on galactolipids, are heterologous fatty acid desaturases, even when from the same plant. It should be noted that these fatty acid desaturases have never been isolated and characterized as proteins. Accordingly, the terms such as "delta-12 desaturase" and "delta-15 desaturase" are used as a convenience to describe the proteins encoded by nucleic acid fragments

that have been isolated based on the phenotypic effects caused by their disruption. They do not imply any catalytic mechanism. For example, delta-12 desaturase refers to the enzyme that catalyzes the formation of a double bond between carbons 12 and 13 of an 18 carbon fatty acid irrespective of whether it "counts" the carbons from the methyl, carboxyl end, or the first double bond. The term "fatty acid desaturase-related enzyme" refers to enzymes whose catalytic product may not be a carbon-carbon double bond but whose mechanism of action is similar to that of a fatty acid desaturase (that is, catalysis of the displacement of a carbon-hydrogen bond of a fatty acid chain to form a fatty-hydroxyacyl intermediate or end-product). Examples include delta-12 hydroxylase which means a delta-12 fatty acid hydroxylase or the oleate hydroxylase responsible for the synthesis of ricinoleic acid from oleic acid.

The term "nucleic acid" refers to a large molecule which can be single-stranded or double-stranded, composed of monomers (nucleotides) containing a sugar, a phosphate and either a purine or pyrimidine. A "nucleic acid fragment" is a fraction of a given nucleic acid molecule. In higher plants, deoxyribonucleic acid (DNA) is the genetic material while ribonucleic acid (RNA) is involved in the transfer of the information in DNA into proteins. A "genome" is the entire body of genetic material contained in each cell of an organism. The term "nucleotide sequence" refers to the sequence of DNA or RNA polymers, which can be single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers. The term "oligomer" refers to short nucleotide sequences, usually up to 100 bases long. As used herein, the term "homologous to" refers

to the relatedness between the nucleotide sequence of two nucleic acid molecules or between the amino acid sequences of two protein molecules. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) Nucleic Acid Hybridisation, IRL Press, Oxford, U.K.); or by the comparison of sequence similarity between two nucleic acids or proteins, such as by the method of Needleman et al. (J. Mol. Biol. (1970) 48:443-453). As used herein, "substantially homologous" refers to nucleotide sequences that have more than 90% overall identity at the nucleotide level with the coding region of the claimed sequence, such as genes and pseudo-genes corresponding to the coding regions. The nucleic acid fragments described herein include molecules which comprise possible variations, both man-made and natural, such as but not limited to (a) those that involve base changes that do not cause a change in an encoded amino acid, or (b) which involve base changes that alter an amino acid but do not affect the functional properties of the protein encoded by the DNA sequence, (c) those derived from deletions, rearrangements, amplifications, random or controlled mutagenesis of the nucleic acid fragment, and (d) even occasional nucleotide sequencing errors.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding) and following (3' non-coding) the coding region. "Fatty acid desaturase gene" refers to a nucleic acid fragment that expresses a protein with fatty acid desaturase activity. "Native" gene refers to an isolated gene with its own regulatory sequences as found in nature. "Chimeric gene" refers to a gene that comprises heterogeneous regulatory and

coding sequences not found in nature. "Endogenous" gene refers to the native gen normally found in its natural location in the genome and is not isolated. A "foreign" gene refers to a gene not normally found in the host organism but that is introduced by gene transfer. "Pseudo-gene" refers to a genomic nucleotide sequence that does not encode a functional enzyme.

"Coding sequence" refers to a DNA sequence that codes for a specific protein and excludes the non-coding sequences. It may constitute an "uninterrupted coding sequence", i.e., lacking an intron or it may include one or more introns bounded by appropriate splice junctions. An "intron" is a nucleotide sequence that is transcribed in the primary transcript but that is removed through cleavage and re-ligation of the RNA within the cell to create the mature mRNA that can be translated into a protein.

"Initiation codon" and "termination codon" refer to a unit of three adjacent nucleotides in a coding sequence that specifies initiation and chain termination, respectively, of protein synthesis (mRNA translation). "Open reading frame" refers to the coding sequence uninterrupted by introns between initiation and termination codons that encodes an amino acid sequence.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA

refers to RNA transcript that includes the mRNA. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene by interfering with the processing, transport and/or translation of its primary transcript or mRNA. The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. In addition, as used herein, antisense RNA may contain regions of ribozyme sequences that increase the efficacy of antisense RNA to block gene expression. "Ribozyme" refers to a catalytic RNA and includes sequence-specific endoribonucleases.

As used herein, "suitable regulatory sequences" refer to nucleotide sequences in native or chimeric genes that are located upstream (5'), within, and/or downstream (3') to the nucleic acid fragments of the invention, which control the expression of the nucleic acid fragments of the invention. The term "expression", as used herein, refers to the transcription and stable accumulation of the sense (mRNA) or the antisense RNA derived from the nucleic acid fragment(s) of the invention that, in conjunction with the protein apparatus of the cell, results in altered levels of the fatty acid desaturase(s). Expression or overexpression of the gene involves transcription of the gene and translation of the mRNA into precursor or mature fatty acid desaturase proteins. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of preventing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Cosuppression" refers to the expression of a foreign

gene which has substantial homology to an endog nous gene resulting in the suppression of expression of both the foreign and the endogenous gene. "Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Promoter" refers to a DNA sequence in a gene, usually upstream (5') to its coding sequence, which controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. In artificial DNA constructs promoters can also be used to transcribe antisense RNA. Promoters may also contain DNA sequences that are involved in the binding of protein factors which control the effectiveness of transcription initiation in response to physiological or developmental conditions. It may also contain enhancer elements. An "enhancer" is a DNA sequence which can stimulate promoter activity. It may be an innate element of the promoter or a heterologous element inserted to enhance the level and/or tissue-specificity of a promoter. "Constitutive promoters" refers to those that direct gene expression in all tissues and at all times. "Tissue-specific" or "development-specific" promoters as referred to herein are those that direct gene expression almost exclusively in specific tissues, such as leaves or seeds, or at specific development stages in a tissue, such as in early or late embryogenesis, respectively.

The "3' non-coding sequences" refers to the DNA sequence portion of a gene that contains a polyadenylation signal and any other regulatory signal capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by

affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor.

"Transformation" herein refers to the transfer of a foreign gene into the genome of a host organism and its genetically stable inheritance. "Restriction fragment length polymorphism" (RFLP) refers to different sized restriction fragment lengths due to altered nucleotide sequences in or around variant forms of genes. "Molecular breeding" refers to the use of DNA-based diagnostics, such as RFLP, RAPDs, and PCR in breeding. "Fertile" refers to plants that are able to propagate sexually.

"Plants" refer to photosynthetic organisms, both eukaryotic and prokaryotic, whereas the term "Higher plants" refers to eukaryotic plants. "Oil-producing species" herein refers to plant species which produce and store triacylglycerol in specific organs, primarily in seeds. Such species include soybean (Glycine max), rapeseed and canola (including Brassica napus, B. campestris), sunflower (Helianthus annus), cotton (Gossypium hirsutum), corn (Zea mays), cocoa (Theobroma cacao), safflower (Carthamus tinctorius), oil palm (Elaeis guineensis), coconut palm (Cocos nucifera), flax (Linum usitatissimum), castor (Ricinus communis) and peanut (Arachis hypogaea). The group also includes non-agronomic species which are useful in developing appropriate expression vectors such as tobacco, rapid cycling Brassica species, and Arabidopsis thaliana, and wild species which may be a source of unique fatty acids.

"Sequence-dependent protocols" refer to techniques that rely on a nucleotide sequence for their utility. Examples of sequence-dependent protocols include, but are not limited to, the methods of nucleic acid and oligomer hybridization and methods of DNA and RNA

amplification such as are exemplified in various uses of the polymerase chain reaction (PCR).

Various solutions used in the experimental manipulations are referred to by their common names such as "SSC", "SSPE", "Denhardt's solution", etc. The composition of these solutions may be found by reference to Appendix B of Sambrook, et al. (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press).

T-DNA Mutagenesis and Identification of an Arabidopsis Mutant Defective in Microsomal Delta-12 Desaturation

In T-DNA mutagenesis (Feldmann, et al., Science (1989) 243:1351-1354), the integration of T-DNA in the genome can interrupt normal expression of the gene at or near the site of the integration. If the resultant mutant phenotype can be detected and shown genetically to be tightly linked to the T-DNA insertion, then the "tagged" mutant locus and its wild type counterpart can be readily isolated by molecular cloning by one skilled in the art.

Arabidopsis thaliana seeds were transformed by Agrobacterium tumefaciens C58C1rif strain harboring the avirulent Ti-plasmid pGV3850::pAK1003 that has the T-DNA region between the left and right T-DNA borders replaced by the origin of replication region and ampicillin resistance gene of plasmid pBR322, a bacterial kanamycin resistance gene, and a plant kanamycin resistance gene (Feldmann, et al., Mol. Gen. Genetics (1987) 208:1-9). Plants from the treated seeds were self-fertilized and the resultant progeny seeds, germinated in the presence of kanamycin, were self-fertilized to give rise to a population, designated T3, that was segregating for T-DNA insertions. T3 seeds from approximately 1700 T2 plants were germinated and grown under controlled

environment. One leaf from each of ten T3 plants of each line were pooled and analyzed for fatty acid composition. One line, designated 658, showed an incresed level of oleic acid (18:1). Analysis of twelve individual T3 seeds of line 658 identified two seeds that contained greater than 36% oleic acid while the remaining seeds contained 12-22% oleic acid. The mutant phenotype of increased level of oleic acid in leaf and seed tissues of line 658 and its segregation in individual T3 seeds suggested that line 658 harbors a mutation that affects desaturation of oleic acid to linoleic acid in both leaf and seed tissues. When approximately 200 T3 seeds of line 658 were tested for their ability to germinate in the presence of kanamycin, four kanamycin-sensitive seeds were identified, suggesting multiple, possibly three, T-DNA inserts in the original T2 line. When progeny seeds of 100 individual T3 plants were analyzed for fatty acid composition and their ability to germinate on kanamycin, one plant, designated 658-75, was identified whose progeny segregated 7 wild type:2 mutant for the increased oleic acid and 28 sensitive:60 resistant for kanamycin resistance. Approximately 400 T4 progeny seeds of derivative line 658-75 were grown and their leaves analyzed for fatty acid composition. Ninety one of these seedlings were identified as homozygous for the mutant (high oleic acid) phenotype. Eighty-three of these homozygous plants were tested for the presence of nopaline, another marker for T-DNA, and all of them were nopaline positive. On the basis of these genetic studies it was concluded that the mutation in microsomal delta-12 desaturation was linked to the T-DNA.

Isolation of Arabidopsis 658-75 Genomic DNA Containing the Disrupted Gene Controlling Microsomal Delta-12 Desaturation

In order to isolate the gene controlling microsomal delta-12 desaturation from wild-type *Arabidopsis*, a T-DNA-plant DNA "junction" fragment containing a T-DNA border integrated into the host plant DNA was isolated from the homozygous mutant plants of the 658-75 line of *Arabidopsis*. For this, genomic DNA from the mutant plant was isolated and completely digested by either Bam HI or Sal I restriction enzymes. In each case, one of the resultant fragments was expected to contain the origin of replication and ampicillin-resistance gene of pBR322 as well as the left T-DNA-plant DNA junction fragment. Such fragments were rescued as plasmids by ligating the digested genomic DNA fragments at a dilute concentration to facilitate self-ligation and then using the ligated fragments to transform *E. coli* cells. While no ampicillin-resistant colony was obtained from the plasmid rescue of Sal I-digested plant genomic DNA, a single ampicillin-resistant colony was obtained from the plasmid rescue of Bam HI-digested plant genomic DNA. The plasmid obtained from this transformant was designated p658-1. Restriction analysis of plasmid p658-1 with Bam HI, Sal I and Eco RI restriction enzymes strongly suggested that it contained the expected 14.2 kb portion of the T-DNA (containing pBR322 sequences) and a putative plant DNA/left T-DNA border fragment in a 1.6 kB Eco RI-Bam HI fragment. The 1.6 kb Eco RI-Bam HI fragment was subcloned into pBluescript SK [Stratagene] by standard cloning procedures described in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press) and the resultant plasmid, designated pS1658.

Isolation of Microsomal Delta-12 Desaturase cDNA and Gene from Wild type Arabidopsis

The 1.6 kb Eco RI-Bam HI fragment, which contained the putative plant DNA flanking T-DNA, in plasmid p658-1 was isolated and used as a radiolabeled hybridization probe to screen a cDNA library made to polyA$^+$ mRNA from the above-ground parts of Arabidopsis thaliana plants, which varied in size from those that had just opened their primary leaves to plants which had bolted and were flowering [Elledge et al. (1991) Proc. Natl. Acad Sci. USA 88:1731-1735]. The cDNA inserts in the library were made into an Xho I site flanked by Eco RI sites in lambda Yes vector [Elledge et al. (1991) Proc. Natl. Acad Sci. USA 88:1731-1735]. Of the several positively-hybridizing plaques, four were subjected to plaque purification. Plasmids were excised from the purified phages by site-specific recombination using the cre-lox recombination system in E. coli strain BNN132 [Elledge et al. (1991) Proc. Natl. Acad Sci. USA 88:1731-1735]. The four excised plasmids were digested by Eco RI restriction enzyme and shown to contain cDNA inserts ranging in size between 1 kB and 1.5 kB. Partial nucleotide sequence determination and restriction enzyme mapping of all four cDNAs revealed their common identity.

The partial nucleotide sequences of two cDNAs, designated pSF2b and p92103, containing inserts of ca. 1.2 kB and ca. 1.4 kB, respectively, were determined. The composite sequence derived from these plasmids is shown as SEQ ID NO:1 and is expected to be contained completely in plasmid p92103. SEQ ID NO:1 shows the 5' to 3' nucleotide sequence of 1372 base pairs of the Arabidopsis cDNA which encodes microsomal delta-12 fatty acid desaturase. Nucleotides 93-95 are the putative initiation codon of the open reading frame (nucleotides

93-1244), (identified by comparison of other plant delta-12 desaturases in this application). Nucleotides 1242-1244 are the termination codon. Nucleotides 1 to 92 and 1245-1372 are the 5' and 3' untranslated nucleotides, respectively. The 383 amino acid protein sequence in SEQ ID NO:2 is that deduced from the open reading frame and has an estimated molecular weight of 44 kD.

The gene corresponding to SEQ ID NO:1 was isolated by screening an Arabidopsis genomic DNA library using radiolabeled pSF2b cDNA insert, purifying the positively-hybridizing plaque, and subcloning a 6 kB Hind III insert fragment from the phage DNA in pBluescript vector. The sequence of 2973 nucleotides of the gene is shown in SEQ ID NO:15. Comparison of the sequences of the gene (SEQ ID NO:15) and the cDNA (SEQ ID NO:1) revealed the presence of a single intron of 1134 bp at a position between nucleotide positions 88 and 89 of the cDNA, which is 4 nucleotides 5' to the initiation codon.

The 1.6 kB Eco RI-Bam HI genomic border fragment insert in pS1658 was also partially sequenced from the Bam HI and Eco RI ends. Comparison of the nucleotide sequences of the gene (SEQ ID NO:15), the cDNA (SEQ ID NO:1), the border fragment, and the published sequence of the left end of T-DNA (Yadav et al., (1982) Proc. Natl. Acad. Sci. 79:6322-6326) revealed that a) the sequence of the first 451 nucleotides of the border fragment from the Bam HI end is collinear with that of nucleotides 539 (Bam HI site) to 89 of the cDNA, b) from the Eco RI end, the border fragment is collinear from nucleotides 1 to 61 with that of the left end of T-DNA (except for a deletion of 9 contiguous nucleotides at position 42 in the border fragment), and is collinear from nucleotides 57 to 104 with that of nucleotides

41-88 of the cDNA, and c) the sequence divergences between the border fragment and the cDNA are due to the presence of the intron in the border fragment. These results show that the T-DNA disrupted the microsomal delta-12 desaturase gene in the transcribed region between the promoter and the coding region and 5' to the intron in the untranslated sequence.

A phage DNA containing Arabidopsis microsomal delta-12 desaturase gene was used as a RFLP marker on a Southern blot containing genomic DNA from several progeny of Arabidopsis thaliana (ecotype Wassileskija and marker line W100 ecotype Landesberg background) digested with Hind III. This mapped the microsomal delta-12 desaturase gene 13.6 cM proximal to locus c3838, 9.2 cM distal to locus 1At228, and 4.9 cM proximal to Fad D locus on chromosome 3 [Koorneef, M. et al., (1993) in Genetic Maps, Ed. O'Brien, S. J.; Yadav et al. (1993) Plant Physiology 103:467-476]. This position corresponds closely to previously suggested locus for microsomal delta-12 desaturation (*Fad 2*) [Hugly, S. et al., (1991) Heredity 82:4321].

The open reading frames in SEQ ID NO:1 and in sequences encoding Arabidopsis microsomal delta-15 desaturase [WO 9311245], Arabidopsis plastid delta-15 desaturase [WO 9311245], and cyanobacterial desaturase, *des A*, [Wada, et al., Nature (1990) 347:200-203; Genbank ID:CSDESA; GenBank Accession No:X53508] as well as their deduced amino acid sequences were compared by the method of Needleman et al. [J. Mol. Biol. (1970) 48:443-453] using gap weight and gap length weight values of 5.0 and 0.3, respectively, for the nucleotide sequences and 3.0 and 0.1, respectively, for protein sequences. The overall identities are summarized in Table 2.

TABLE 2

Percent Identity Between Different Fatty Acid Desaturases at the Nucleotide and Amino Acid Levels

		a3	ad	des A
a2	nucleotide	48(8 gaps)	46(6 gaps)	43(10 gaps)
	amino acid	39(9 gaps)	34(8 gaps)	24(10 gaps)
a3	nucleotide	-	65(1 gap)	43(9 gaps)
	amino acid	-	65(2 gaps)	26(11 gaps)
ad	nucleotide	-	-	43(9 gaps)
	amino acid	-	-	26(11 gaps)

a2, a3, ad, and *des A* refer, respectively, to SEQ ID NO:1/2, Arabidopsis microsomal delta-15 desaturase, Arabidopsis plastid delta-15 desaturase, and cyanobacterial desaturase, *des A*. The percent identities in each comparison are shown at both the nucleotide and amino acid levels; the number of gaps imposed by the comparisons are shown in brackets following the percent identities. As expected on the basis of unsuccessful attempts in using delta-15 fatty acid nucleotide sequences as hybridization probes to isolate nucleotide sequences encoding microsomal delta-12 fatty acid desaturase, the overall homology at the nucleotide level between microsomal delta-12 fatty acid desaturase (SEQ ID NO:1) and the nucleotide sequences encoding the other three desaturases is poor (ranging between 43% and 48%). At the amino acid level too, the microsomal delta-12 fatty acid desaturase (SEQ ID NO:2) is poorly related to cyanobacterial *des A* (less than 24% identity) and the plant delta-15 desaturases (less than 39% identity).

While the overall relatedness between the deduced amino acid sequence of the said invention and the published fatty acid desaturases is limited, more significant identities are observed in shorter stretches

of amino acid sequences in the above comparisons. These results confirm d that the T-DNA in line 658-75 had interrupted the normal expression of a fatty acid desaturase gene. Based on the fatty acid phenotype of homozygous mutant line 658-75, Applicants concluded that SEQ ID NO:1 encoded the delta-12 desaturase. Further, Applicants concluded that it was the microsomal delta-12 desaturase, and not the chloroplastic delta-12 desaturase, since: a) the mutant phenotype was expressed strongly in the seed but expressed poorly, if at all, in the leaf of line 658-75, and b) the delta-12 desaturase polypeptide, by comparison to the microsomal and plastid delta-15 desaturase polypeptides [WO 9311245], did not have an N-terminal extension of a transit peptide expected for a nuclear-encoded plastid desaturase.

Plasmid p92103 was deposited on October 16, 1992 with the American Type Culture Collection of Rockville, Maryland under the provisions of the Budapest Treaty and bears accession number ATCC 69095.

<u>Expression Of Microsomal Delta-12 Fatty Acid Desaturase In Arabidopsis *Fad2-1* Mutant To Complement Its Mutation In Delta-12 Fatty Acid Desaturation</u>

To confirm the identity of SEQ ID NO:1 (<u>Arabidopsis</u> microsomal delta-12 fatty acid desaturase cDNA) a chimeric gene comprising of SEQ ID NO:1 was transformed into an <u>Arabidopsis</u> mutant affected in microsomal delta-12 fatty acid desaturation. For this, the ca. 1.4 kb Eco RI fragment containing the cDNA (SEQ ID NO:1) was isolated from plasmid p92103 and sub-cloned in pGA748 vector [An et. al.(1988) Binary Vectors. In: Plant Molecular Biology Manual. Eds Gelvin, S. B. et al. Kluwer Academic Press], which was previously linearized with Eco RI restriction enzyme. In one of the resultant binary plasmid, designated pGA-Fad2, the cDNA was placed

in the sense orientation behind the CaMV 35S promotor of the vector to provide constitutive expression.

Binary vector pGA-Fad2 was transformed by the freeze/thaw method [Holsters et al. (1978) Mol. Gen. Genet. 163:181-187] into Agrobacterium tumefaciens strain R1000, carrying the Ri plasmid pRiA4b from Agrobacterium rhizogenes [Moore et al., (1979) Plasmid 2:617-626] to result in transformants R1000/pGA-Fad2.

Agrobacterium strains R1000 and R1000/pGA-Fad2 were used to transform Arabidopsis mutant fad2-1 [Miquel, M. & Browse, J. (1992) Journal of Biological Chemistry 267:1502-1509] and strain R1000 was used to transform wild type Arabidopsis. Young bolts of plants were sterilized and cut so that a single node was present in each explant. Explants were inoculated by Agrobacteria and incubated at 25°C in the dark on drug-free MS minimal organics medium with 30 g/L sucrose (Gibco). After four days, the explants were transferred to fresh MS medium containing 500 mg/L cefotaxime and 250 mg/ml carbenicillin for the counterselection of Agrobacterium. After 5 days, hairy roots derived from R1000/pGA-Fad2 transformation were excised and transferred to the same medium containing 50 mg/ml kanamycin. Fatty acid methyl esters were prepared from 5-10 mm of the roots essentially as described by Browse et al., (Anal. Biochem. (1986) 152:141-145) except that 2.5% H_2SO_4 in methanol was used as the methylation reagent and samples were heated for 1.5 h at 80°C to effect the methanolysis of the seed triglycerides. The results are shown in Table 3. Root samples 41 to 46, 48 to 51, 58, and 59 are derived from transformation of fad2-1 plants with R1000/pGA Fad2; root samples 52, 53, and 57 were derived from transformation of fad2-1 plants with R1000 and serve as controls; root sample 60 is derived from

transformation of wild type Arabidopsis with R1000 and also serves as a control.

TABLE 3

Fatty acid Composition in Transgenic Arabidopsis fad2-1 Hairy Roots Transformed with Agrobacterium R1000/pGA-fad2

Sample	16:0	16:1	18:0	18:1	18:2	18:3
41	24.4	1.8	1.7	5.0	29.4	33.8
42	25.6	3.7	1.3	20.0	22.0	27.5
43	23.6	-	1.6	7.2	27.6	36.1
44	24.4	1.3	4.6	16.0	18.1	33.6
45	20.7	-	8.1	44.7	11.8	14.8
46	20.1	-	1.8	7.5	33.7	36.0
48	26.1	2.9	2.1	9.5	17.6	33.4
49	30.8	1.0	2.4	8.7	18.7	31.1
50	19.8	1.9	3.3	27.7	21.8	24.4
51	20.9	1.1	5.0	13.7	25.0	32.1
58	23.5	0.3	1.4	3.6	22.1	45.9
59	22.6	0.6	1.4	2.8	29.9	40.4
52, cont.	12.3	-	2.6	64.2	4.6	16.4
53, cont.	20.3	9.1	2.2	55.2	1.7	9.2
57, cont.	10.4	2.4	0.7	65.9	3.8	12.7
60,WT	23.0	1.7	0.8	6.0	35.0	31.8

These results show that expression of Arabidopsis microsomal delta-12 desaturase in a mutant Arabidopsis lacking delta-12 desaturation can result in partial to complete complementation of the mutant. The decrease in oleic acid levels in transgenic roots is accompanied by increases in the levels of both 18:2 and 18:3. Thus, overexpression of this gene in other oil crops, especially canola, which is a close relative of Arabidopsis and which naturally has high levels of 18:1 in seeds, is also expected to result in higher levels of 18:2, which in conjunction with

the overexpression of the microsomal delta-15 fatty acid desaturase will result in very high levels of 18:3.

Using Arabidopsis Microsomal Delta-12 Desaturase cDNA as a Hybridization Probe to Isolate Microsomal Delta-12 Desaturase cDNAs from Other Plant Species

Evidence for conservation of the delta-12 desaturase sequences amongst species was provided by using the Arabidopsis cDNA insert from pSF2b as a hybridization probe to clone related sequences from Brassica napus, and soybean. Furthermore, corn and castor bean microsomal delta-12 fatty acid desaturase were isolated by PCR using primers made to conserved regions of microsomal delta-12 desaturases.

Cloning of a Brassica napus Seed cDNA Encoding Seed Microsomal Delta-12 Fatty Acid Desaturase

For the purpose of cloning the Brassica napus seed cDNA encoding a delta-12 fatty acid desaturase, the cDNA insert from pSF2b was isolated by digestion of pSF2b with EcoR I followed by purification of the 1.2 kb insert by gel electrophoresis. The 1.2 kb fragment was radiolabeled and used as a hybridization probe to screen a lambda phage cDNA library made with poly A^+ mRNA from developing Brassica napus seeds 20-21 days after pollination. Approximately 600,000 plaques were screened under low stringency hybridization conditions (50 mM Tris pH 7.6, 6X SSC, 5X Denhardt's, 0.5% SDS, 100 ug denatured calf thymus DNA and 50°C) and washes (two washes with 2X SSC, 0.5% SDS at room temperature for 15 min each, then twice with 0.2X SSC, 0.5% SDS at room temperature for 15 min each, and then twice with 0.2X SSC, 0.5% SDS at 50°C for 15 min each). Ten strongly-hybridizing phage were plaque-purified and the size of the cDNA inserts was determined by PCR amplication of the insert using phage as template and

T3/T7 oligomers for primers. Two of these phages, 165D and 165F, had PCR amplified inserts of 1.6 kb and 1.2 kb respectively and these phages were also used to excise the phagemids as described above. The phagemid derived from phage 165D, designated pCF2-165D, contained a 1.5 kb insert which was sequenced completely on one strand. SEQ ID NO:3 shows the 5' to 3' nucleotide sequence of 1394 base pairs of the Brassica napus cDNA which encodes delta-12 desaturase in plasmid pCF2-165d. Nucleotides 99 to 101 and nucleotides 1248 to 1250 are, respectively, the putative initiation codon and the termination codon of the open reading frame (nucleotides 99 to 1250). Nucleotides 1 to 98 and 1251 to 1394 are, respectively, the 5' and 3' untranslated nucleotides. The 383 amino acid protein sequence deduced from the open reading frame in SEQ ID NO:3 is shown in SEQ ID NO:4. While the other strand can easily be sequenced for confirmation, comparisons of SEQ ID NOS:1 and 3 and of SEQ ID NOS:2 and 4, even admitting of possible sequencing errors, showed an overall homology of approximately 84% at both the nucleotide and amino acid levels, which confirmed that pCF2-165D is a Brassica napus seed cDNA that encoded delta-12 desaturase. Plasmid pCF2-165D has been deposited on October 16, 1992 with the American Type Culture Collection of Rockville, Maryland under the provisions of the Budapest Treaty and bears accession number ATCC 69094.

Cloning of a Soybean (Glycine max) cDNA Encoding Seed Microsomal Delta-12 Fatty Acid Desaturase

A cDNA library was made to poly A^+ mRNA isolated from developing soybean seeds, and screened as described above. Radiolabelled probe prepared from pSF2b as described above was added, and allowed to hybridize for 18 h at 50°C. The probes were washed as described

above. Autoradiography of the filters indicated that there were 14 strongly hybridizing plaques, and numerous weakly hybridizing plaques. Six of the 14 strongly hybridizing plaques were plaque purified as described above and the cDNA insert size was determined by PCR amplication of the insert using phage as template and T3/T7 oligomers for primers. One of these phages, 169K, had an insert sizes of 1.5 kb and this phage was also used to excise the phagemid as described above. The phagemid derived from phage 169K, designated pSF2-169K, contained a 1.5 kb insert which was sequenced completely on both strands. SEQ ID NO:5 shows the 5' to 3' nucleotide sequence of 1473 base pairs of soybean (Glycine max) cDNA which encodes delta-12 desaturase in plasmid pSF2-169K. Nucleotides 108 to 110 and nucleotides 1245 to 1247 are, respectively, the putative initiation codon and the termination codon of the open reading frame (nucleotides 108 to 1247). Nucleotides 1 to 107 and 1248 to 1462 are, respectively, the 5' and 3' untranslated nucleotides. The 380 amino acid protein sequence deduced from the open reading frame in SEQ ID NO:5 is shown in SEQ ID NO:6. Comparisons of SEQ ID NOS:1 and 5 and of SEQ ID NOS:2 and 6, even admitting of possible sequencing errors, showed an overall homology of approximately 65% at the nucleotide level and approximately 70% at the amino acid level, which confirmed that pSF2-169K is a soybean seed cDNA that encoded delta-12 desaturase. A further description of this clone can be obtained by comparison of the SEQ ID NO:1, SEQ ID NO:3, and SEQ ID NO:5 and by analyzing the phenotype of transgenic plants produced by using chimeric genes incorporating the insert of pSF2-169K, in sense or antisense orientation, with suitable regulatory sequences. Plasmid pSF2-169K was deposited on October 16, 1992 with the American Type Culture

Collection of Rockville, Maryland under the provisions of the Budapest Treaty and bears accession number ATCC 69092.

Cloning of a Corn (Zea mays) cDNA Encoding Seed Microsomal Delta-12 Fatty Acid Desaturase

Corn microsomal delta-12 desaturase cDNA was isolated using a PCR approach. For this, a cDNA library was made to poly A^+ RNA from developing corn embryos in Lambda Zap II vector. This library was used as template for PCR using sets of degenerate oligomers NS3 (SEQ ID NO:13) and RB5A/B (SEQ ID NOS:16 and 17) as sense and antisense primers, respectively. NS3 and RB5A/B correspond to stretches of amino acids 101-109 and 318-326, respectively, of SEQ ID NO:2, which are conserved in most microsomal delta-12 desaturases (for example, SEQ ID NOS:2, 4, 6, 8). PCR was carried out using a PCR kit (Perkin-Elmer) by 40 cycles of 94°C 1', 45°C, 1 min, and 55°C, 2 min. Analyses of the PCR products on an agarose gel showed the presence of a product of the expected size (720 bp), which was absent in control reactions containing either the sense or antisense primers alone. The fragment was gel purified and then used as a probe for screening the corn cDNA library at 60°C as described above. One positively-hybridizing plaque was purified and partial sequence determination of its cDNA showed it to be a nucleotide sequence encoding microsomal delta-12 desaturase but truncated at the 3' end. The cDNA insert encoding the partial desaturase was gel isolated and used to probe the corn cDNA library again. Several positive plaques were recovered and characterized. DNA sequence analysis revealed that all of these clones seem to represent the same sequence with the different length of 5' or 3' ends. The clone containing the longest insert,

designated pFad2#1, was sequenced completely. The total length of the cDNA is 1790 bp (SEQ ID NO:7) comprising of an open reading frame from nucleotide 165 to 1328 bp that encoded a polypeptide of 388 amino acids. The deduced amino acid sequence of the polypeptide (SEQ ID NO:8) shared overall identities of 71%, 40%, and 38% to Arabidopsis microsomal delta-12 desaturase, Arabidopsis microsomal delta-15 desaturase, and Arabidopsis plastid delta-15 desaturase, respectively. Furthermore, it lacked an N-terminal amino acid extension that would indicate it is a plastid enzyme. Based on these considerations, it is concluded that it encodes a microsomal delta-12 desaturase.

Isolation of cDNAs Encoding Delta-12 Microsomal Fatty Acid Desaturases and Desaturase-Related Enzymes from Castor Bean Seed

Polysomal mRNA was isolated from castor beans of stages I-II (5-10 DAP) and also from castor beans of stages IV-V (20-25 DAP). Ten ng of each mRNA was used for separate RT-PCR reactions, using the Perkin-Elmer RT-PCR kit. The reverse transcriptase reaction was primed with random hexamers and the PCR reaction with degenerate delta-12 desaturase primers NS3 and NS9 (SEQ ID NOS:13 and 14). The annealing-extension temperature of the PCR reaction was 50°C. A DNA fragment of approx. 700 bp was amplified from both stage I-II and stage IV-V mRNA. The amplified DNA fragment from one of the reactions was gel purified and cloned into a pGEM-T vector using the Promega pGEM-T PCR cloning kit to create the plasmid pRF2-1C. The 700 bp insert in pRF2-1C was sequenced, as described above, and the resulting DNA sequence is shown in SEQ ID NO:9. The DNA sequence in SEQ ID NO:9 contains an open-reading frame encoding 219 amino acids (SEQ ID NO:10) which has 81% identity (90% similarity) with amino acids 135 to 353 of

the *Arabidopsis* microsomal delta-12 desaturase described in SEQ ID NO:2. The cDNA insert in pRF2-1C is therefore a 676 bp fragment of a full-length cDNA encoding a castor bean seed microsomal delta-12 desaturase. The full length castor bean seed microsomal delta-12 desaturase cDNA may isolated by screening a castor seed cDNA library, at 60°C, with the labeled insert of pRF2-1C as described in the example above. The insert in pRF2-1C may also be used to screen castor bean libraries at lower temperatures to isolate delta-12 desaturase-related sequences, such as the delta-12 hydroxylase.

A cDNA library made to poly A^+ mRNA isolated from developing castor beans (stages IV-V, 20-25 DAP) was screened as described above. Radiolabeled probe prepared from pSF2b or pRF2-1C, as described above, were added, and allowed to hybridize for 18 h at 50°C. The filters were washed as described above. Autoradiography of the filters indicated that there were numerous hybridizing plaques, which appeared either strongly-hybridising or weakly-hybridising. Three of the strongly hybridisng plaques (190A-41, 190A-42 and 190A-44) and three of the weakly hybridising plaques, (190B-41, 190b-43 and 197c-42), were plaque purified using the methods described above. The cDNA insert size of the purified phages were determined by PCR amplication of the insert using phage as template and lambda-gt11 oligomers (Clontech lambda-gt11 Amplimers) for primers. The PCR-amplified inserts of the amplified phages were subcloned into pBluescript (Pharmacia) which had been cut with Eco RI and filled in with Klenow (Sambrook et al. (Molecular Cloning, A Laboratory Approach, 2nd. ed. (1989) Cold Spring Harbor Laboratory Press). The resulting plasmids were called pRF190a-41, pRF190a-42, pRF190a-44, pRF190b-41, pRF190b-43 and

pRF197c-42. All of the inserts were about 1.1 kb with the exception of pRF197c-42 which was approx. 1.5 kb. The inserts in the plasmids were sequenced as described above. The insert in pRF190b-43 did not contain any open reading frame and was not identified. The inserts in pRF190a-41, pRF190a-42, pRF190a-44 and pRF190b-41 were identical. The insert in pRF197c-42 contained all of the nucleotides of the inserts in pRF190a-41, pRF190a-42, pRF190a-44 and pRF190b-41 plus an additional approx. 400 bp. It was deduced therefore that the insert in pRF197c-42 was a longer version of the inserts in pRF190a-41, pRF190a-42, pRF190a-44 and pRF190b-41 and all were derived from the same full-length mRNA. The complete cDNA sequence of the insert in plasmid pRF197c-42 is shown in SEQ ID NO:11. The deduced amino acid sequence of SEQ ID NO:11, shown in SEQ ID NO:12, is 78.5% identical (90% similarity) to the castor microsomal delta-12 desaturase described above (SEQ ID NO:10) and 66% identical (80% similarity) to the Arabidopsis delta-12 desaturase amino acid sequence in SEQ ID NO:2. These similarities confirm that pRF197c-42 is a castor bean seed cDNA that encodes a microsomal delta-12 desaturase or a microsomal delta-12 desaturase-related enzyme, such as a delta-12 hydroxylase. Specific PCR primers for pRF2-1C and pRF197c-42 were made. For pRF2-1c the upstream primer was bases 180 to 197 of the cDNA sequence in SEQ ID NO:9. For pRF197c-42 the upstream primer was bases 717 to 743 of the cDNA sequence in SEQ ID NO:11. A common downstream primer was made corresponding to the exact complement of the nucleotides 463 to 478 of the sequence described in SEQ ID NO:9. Using RT-PCR with random hexamers and the above primers it was observed that the cDNA contained in pRF2-1C is expressed in both Stage I-II and Stage IV-V castor bean seeds whereas the cDNA contained in plasmid

pRF197c-42 is expressed only in Stage IV-V castor b an seeds, i.e., it is only expressed in tissue actively synthesizing ricinoleic acid. Thus, it is possible that this cDNA encodes a delta-12 hydroxylase.

There is enough deduced amino acid sequence from the two castor sequences described in SEQ ID NOS:10 and 12 to compare the highly conserved region corresponding to amino acids 311 to 353 of SEQ ID NO:2. When SEQ ID NOS:2, 4, 6, 8, and 10 are aligned by the Hein method described above the consensus sequence corresponds exactly to the amino acids 311 to 353 of SEQ ID NO:2. All of the seed microsomal delta-12 desaturases described above have a high degree of identity with the consensus over this region, namely Arabidopsis (100% identity), soybean (90% identity), corn (95% identity), canola (93% identity) and one (pRF2-1c) of the castor bean sequences (100% identity). The other castor bean seed delta-12 desaturase or desaturase-related sequence (pRF197c-42) however has less identity with the consensus, namely 81% for the deduced amino acid sequence of the insert in pRF197c-42 (described in SEQ ID NO:12). Thus while it remains possible that the insert in pRF197c-42 encodes a microsomal delta-12 desaturase, this observation supports the hypothesis that it encodes a desaturase-related sequence, namely the delta-12 hydroxylase.

An additional approach to cloning a castor bean seed delta-12 hydroxylase is the screening of a differential population of cDNAs. A lambda-Zap (Stratagene) cDNA library made to polysomal mRNA isolated from developing castor bean endosperm (stages IV-V, 20-25 DAP) was screened with ^{32}P-labeled cDNA made from polysomal mRNA isolated from developing castor bean endosperm (stage I-II, 5-10 DAP) and with ^{32}P-labeled cDNA made from polysomal mRNA isolated from developing

castor bean endosperm (stages IV-V, 20-25 DAP). The library was scre ned at a density of 2000 plaques per 137 mm plate so that individual plaques were isolated. About 60,000 plaques were screened and plaques which hybridised with late (stage IV/V) cDNA but not early (stage I/II) cDNA, which corresponded to about 1 in every 200 plaques, were pooled.

The library of differentially expressed cDNAs may be screened with the castor delta-12 desaturase cDNA described above and/or with degenerate oligonucleotides based on sequences of amino conserved among delta-12 desaturases to isolate related castor cDNAs, including the cDNA encoding the delta-12 oleate hydroxylase enzyme. These regions of amino acid conservation may include, but are not limited to amino acids 101 to 109, 137 to 145, and 318 to 327 of the amino acid sequence described in SEQ ID NO:2 or any of the sequences described in Table 7 below. Examples of such oligomers are SEQ ID NOS:13, 14, 16, and 17. The insert in plasmid pCF2-197c may be cut with Eco RI to remove vector sequences, purified by gel electrophoresis and labeled as described above. Degenerate oligomers based on the above conserved amino acid sequences may be labeled with ^{32}P as described above. The cDNAs cloned from the developing endosperm difference library which do not hybridize with early mRNA probe but do hybridize with late mRNA probe and hybridize with either castor delta-12 desaturase cDNA or with an oligomer based on delta-12 desaturase sequences are likely to be the castor delta-12 hydroxylase. The pBluescript vector containing the putative hydroxylase cDNA can be excised and the inserts directly sequenced, as described above.

Clones such as pRF2-1C and pRF197c-42, and other clones from the differential screening, which, based on their DNA sequence, are less related to castor bean seed

microsomal delta-12 desaturases and are not any of the known fatty-acid desaturases described above or in WO 9311245, may be expressed, for example, in soybean embryos or another suitable plant tissue, or in a microorganism, such as yeast, which does not normally contain ricinoleic acid, using suitable expression vectors and transformation protocols. The presence of novel ricinoleic acid in the transformed tissue(s) expressing the castor cDNA would confirm the identity of the castor cDNA as DNA encoding for an oleate hydroxylase.

Sequence Comparisons Among Seed Microsomal Delta-12 Desaturases

The percent overall identities between coding regions of the full-length nucleotide sequences encoding microsomal delta-12 desaturases was determined by their alignment by the method of Needleman et al. (J. Mol. Biol. (1970) 48:443-453) using gap weight and gap length weight values of 5.0 and 0.3 (Table 4). Here, a2, c2, s2, z2 and *des A* refer, respectively, to the nucleotide sequences encoding microsomal delta-12 fatty acid desaturases from Arabidopsis (SEQ ID NO:1), Brassica napus (SEQ ID NO:3), soybean (SEQ ID NO:5), corn (SEQ ID NO:7), and cyanobacterial *des A*, whereas r2 refers to the microsomal delta-12 desaturase or desaturase-related enzyme from castor bean (SEQ ID NO:12).

TABLE 4

Percent Identity Between the Coding Regions of Nucleotide Sequences Encoding Different Microsomal Delta-12 Fatty Acid Desaturases

	c2	s2	z2	*des A*
a2	84	66	64	43
c2	-	65	62	42
s2	-	-	62	42

The overall relatedness between the deduced amino acid sequences of microsomal delta-12 desaturases or desaturase-related enzymes of the invention (i.e., SEQ ID NOS:2, 4, 6, 8, and 12) determined by their alignment by the method of Needleman et al. (J. Mol. Biol. (1970) 48:443-453) using gap weight and gap length weight values of 3.0 and 0.1, respectively, is shown in Table 5. Here a2, c2, s2, z2 and *des A* refer, respectively, to microsomal delta-12 fatty acid desaturases from Arabidopsis (SEQ ID NO:2), Brassica napus (SEQ ID NO:4), soybean (SEQ ID NO:6), corn (SEQ ID NO:8), and cyanobacterial *des A*, whereas r2 refers to the microsomal desaturase or desaturase-related enzyme from castor bean (SEQ ID NO:12). The relatedness between the enzymes is shown as percent overall identity/percent overall similarity.

TABLE 5

Relatedness Between Different Microsomal Delta-12 Fatty Acid Desaturases

	c2	s2	r2	z2	*des A*
a2	84/89	70/85	66/80	71/83	24/50
c2	-	67/80	63/76	69/79	24/51
s2	-	-	67/83	66/82	23/49
r2	-	-	-	61/78	24/51
z2	-	-	-	-	25/49

The high degree of overall identity (60% or greater) at the the amino acid levels between the Brassica napus, soybean, castor and corn enzymes with that of Arabidopsis microsomal delta-12 desaturase and their lack of an N-terminal extension of a transit peptide expected for a nuclear-encoded chloroplast desaturase leads Applicants to conclude that SEQ ID NOS:4, 6, 8, 10, and 12 encode the microsomal delta-12

desaturases or desaturase-related enzymes. Further confirmation of this identity will come from biological function, that is, by analyzing the phenotype of transgenic plants or other organisms produced by using chimeric genes incorporating the above-mentioned sequences in sense or antisense orientation, with suitable regulatory sequences. Thus, one can isolate cDNAs and genes for homologous fatty acid desaturases from the same or different higher plant species, especially from the oil-producing species. Furthermore, based on these comparisons, the Applicants expect all higher plant microsomal delta-12 desaturases from all higher plants to show an overall identity of 60% or more and to be able to readily isolate homologous fatty acid desaturase sequences using SEQ ID NOS:1, 3, 5, 7, 9, and 11 by sequence-dependent protocols.

The overall percent identity at the amino acid level, using the above alignment method, between selected plant desaturases is illustrated in Table 6.

TABLE 6

Percent Identity Between Selected Plant Fatty Acid Desaturases at the Amino Acid Level

	a3	aD	c3	cD	S3
a2	38	33	38	35	34
a3	-	65	93	66	67
aD	-	-	66	87	65
c3	-	-	-	67	67
cD	-	-	-	-	65

In Table 6, a2, a3, ad, c3, cD, and S3 refer, respectively, to SEQ ID NO:2, *Arabidopsis* microsomal delta-15 desaturase, *Arabidopsis* plastid delta-15 desaturase, canola microsomal delta-15 desaturase, canola plastid delta-15 desaturase, and soybean microsomal delta-15 desaturase. Based on these

comparisons, the delta-15 d saturases, of both microsomal and plastid types, have overall identities of 65% or more at the amino acid level, even when from the same plant species. Based on the above the Applicants expect microsomal delta-12 desaturases from all higher plants to show similar levels of identity (that is, 60% or more identity at the amino acid level) and that SEQ ID NOS:1, 3, 5, 7, and 9 can also be used as hybridization probe to isolate homologous delta-12 desaturase sequences, and possibly for sequences for fatty acid desaturase-related enzymes, such as oleate hydroxylase, that have an overall amino acid homology of 50% or more.

Similar alignments of protein sequences of plant microsomal fatty acid delta-12 desaturases [SEQ ID NOS:2, 4, 6, and 8] and plant delta-15 desaturases [microsomal and plastid delta-15 desaturases from *Arabidopsis* and *Brassica napus*, WO 9311245] allows identification of amino acid sequences conserved between the different desaturases (Table 7).

TABLE 7

Amino Acid Sequences Conserved Between Plant Microsomal Delta-12 Desaturases and Microsomal and Plastid Delta-15 Desaturases

Region	Conserved AA Positions in SEQ ID NO:2	Consensus Conserved AA Sequence in Δ^{12}Desaturases	Consensus Conserved AA Sequence in Δ^{15}Desaturases	Consensus AA Sequence
A	39-44	AIPPHC	AIPKHC	AIP(P/K)HC
B	86-90	WP(L/I)YW	WPLYW	WP(L/I)YW
C	104-109	AHECGH	GHDCGH	(A/G)H(D/E)CGH
D	130-134	LLVPY	ILVPY	(L/I)LVPY
E	137-142	WKYSHR	WRISHR	W(K/R)(Y/I)SHR
F	140-145	SHRRHH	SHRTHH	SHR(R/T)HH
G	269-274	ITYLQ	VTYLH	(I/V)TYL(Q/H)
H	279-282	LPHY	LPWY	LP(H/W)Y

I	289-294	WL(R/K)GAL	YLRGGL	(W/Y)L(R/K)G(A/G)L
J	296-302	TVDRDYG	TLDRDYG	T(V/L)DRDYG
K	314-321	THVAHHLF	THVIHHLF	THV(A/I)HHLF
L	318-327	HHLFSTMPHY	HHLFPQIPHY	HHFL(S/P)(T/Q)(I/M)PHY

Table 7 shows twelve regions of conserved amino acid sequences, designated A-L (column 1), whose positions in SEQ ID NO:2 are shown in column 2. The consensus sequences for these regions in plant delta-12 fatty acid desaturases and plant delta-15 fatty acid desaturases are shown in columns 3 and 4, respectively; amino acids are shown by standard abbreviations, the underlined amino acids are conserved between the delta-12 and the delta-15 desaturases, and amino acids in brackets represent substitutions found at that position. The consensus sequence of these regions are shown in column 5. These short conserved amino acids and their relative positions further confirm that the isolated isolated cDNAs encode a fatty acid desaturase.

Isolation of Nucleotide Sequences Encoding Homologous and Heterologous Fatty acid Desaturases and Desaturase-like Enzymes

Fragments of the instant invention may be used to isolate cDNAs and genes of homologous and heterologous fatty acid desaturases from the same species as the fragments of the invention or from different species. Isolation of homologous genes using sequence-dependent protocols is well-known in the art and Applicants have demonstrated that *Arabidopsis* microsomal delta-12 desaturase cDNA sequence can be used to isolate cDNA for related fatty acid desaturases from *Brassica napus*, soybean, corn and castor bean.

More importantly, one can use the fragments containing SEQ ID NOS:1, 3, 5, 7, and 9 or their

smaller, more conserved regions to isolate novel fatty acid desaturases and fatty acid desaturase-related enzymes.

In a particular embodiment of the present invention, regions of the nucleic acid fragments of the invention that are conserved between different desaturases may be used by one skilled in the art to design a mixture of degenerate oligomers for use in sequence-dependent protocols aimed at isolating nucleic acid fragments encoding homologous or heterologous fatty acid desaturase cDNA's or genes. For example, in the polymerase chain reaction (Innis, et al., Eds, (1990) PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego), two short pieces of the present fragment of the invention can be used to amplify a longer fatty acid desaturase DNA fragment from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleotide sequences with one primer based on the fragment of the invention and the other on either the poly A^+ tail or a vector sequence. These oligomers may be unique sequences or degenerate sequences derived from the nucleic acid fragments of the invention. The longer piece of homologous fatty acid desaturase DNA generated by this method could then be used as a probe for isolating related fatty acid desaturase genes or cDNAs from Arabidopsis or other species and subsequently identified by differential screening with known desaturase sequences and by nucleotide sequence determination. The design of oligomers, including long oligomers using deoxyinosine, and "guessmers" for hybridization or for the polymerase chain reaction are known to one skilled in the art and discussed in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press). Short stretches of

amino acid sequences that are conserved between cyanobacterial delta-12 desaturase (Wada et al., Nature (1990) 347:200-203) and plant delta-15 desaturases [WO 9311245] were previously used to make oligonucleotides that were degenerate and/or used deoxyinosine/s. One set of these oligomers made to a stretch of 12 amino acids conserved between cyanobacterial delta-12 desaturase and higher plant delta-15 desaturases was successful in cloning the plastid delta-12 desaturase cDNAs; these plant desaturases have more than 50% identity to the cyanobacterial delta-12 desaturase. Some of these oligonucleotides were also used as primers to make polymerase chain reaction (PCR) products using poly A^+ RNA from plants. However, none of the oligonucleotides and the PCR products were successful as radiolabeled hybridization probes in isolating nucleotide sequences encoding microsomal delta-12 fatty acid desaturases. Thus, as expected, none of the stretches of four or more amino acids conserved between *Arabidopsis* delta-12 and *Arabidopsis* delta-15 desaturases are identical in the cyanobacterial desaturase, just like none of the stretches of four or more amino acids conserved between *Arabidopsis* delta-15 and the cyanobacterial desaturase are identical in SEQ ID NO:2. Stretches of conserved amino acids between the present invention and delta-15 desaturases now allow for the design of oligomers to be used to isolate sequences encoding other desaturases and desaturase-related enzymes. For example, conserved stretches of amino acids between delta-12 desaturase and delta-15 desaturase, shown in Table 7, are useful in designing long oligomers for hybridization as well as shorter ones for use as primers in the polymerase chain reaction. In this regard, sequences conserved between delta-12 and delta-15 desaturases (shown in Table 7)

would be particularly useful. The consensus sequences will also take into account conservative substitutions known to one skilled in the art, such as Lys/Arg, Glu/Asp, Ile/Val/Leu/Met, Ala/Gly, Gln/Asn, and Ser/Thr. Amino acid sequences as short as four amino acids long can successfully be used in PCR [Nunberg et. al. (1989) Journal of Virology 63:3240-3249]. Amino acid sequences conserved between delta-12 desaturases (SEQ ID NOS:2, 4, 6, 8, and 10) may also be used in sequence-dependent protocols to isolate fatty acid desaturases and fatty acid desaturase-related enzymes expected to be more related to delta-12 desaturases, such as the oleate hydroxylase from castor bean. Particularly useful are conserved sequences in column 3 (Table 7), since they are also conserved well with delta-15 desaturases (column 4, Table 7).

Determining the conserved amino acid sequences from diverse desaturases will also allow one to identify more and better consensus sequences that will further aid in the isolation of novel fatty acid desaturases, including those from non-plant sources such as fungi, algae (including the desaturases involved in the desaturations of the long chain n-3 fatty acids), and even cyano-bacteria, as well as other membrane-associated desaturases from other organisms.

The function of the diverse nucleotide fragments encoding fatty acid desaturases or desaturase-related enzymes that can be isolated using the present invention can be identified by transforming plants with the isolated sequences, linked in sense or antisense orientation to suitable regulatory sequences required for plant expression, and observing the fatty acid phenotype of the resulting transgenic plants. Preferred target plants for the transformation are the same as the source of the isolated nucleotide fragments when the

goal is to obtain inhibition of the corresponding endogenous gene by antisense inhibition or cosuppression. Preferred target plants for use in expression or overexpression of the isolated nucleic acid fragments are wild type plants or plants with known mutations in desaturation reactions, such as the Arabidopsis mutants fadA, fadB, fadC, fadD, fad2, and fad3; mutant flax deficient in delta-15 desaturation; or mutant sunflower deficient in delta-12 desaturation. Alternatively, the function of the isolated nucleic acid fragments can be determined similarly via transformation of other organisms, such as yeast or cyanobacteria, with chimeric genes containing the nucleic acid fragment and suitable regulatory sequences followed by analysis of fatty acid composition and/or enzyme activity.

Overexpression of the Fatty Acid Desaturase Enzymes in Transgenic Species

The nucleic acid fragment(s) of the instant invention encoding functional fatty acid desaturase(s), with suitable regulatory sequences, can be used to overexpress the enzyme(s) in transgenic organisms. An example of such expression or overexpression is demonstrated by transformation of the Arabidopsis mutant lacking oleate desaturation. Such recombinant DNA constructs may include either the native fatty acid desaturase gene or a chimeric fatty acid desaturase gene isolated from the same or a different species as the host organism. For overexpression of fatty acid desaturase(s), it is preferable that the introduced gene be from a different species to reduce the likelihood of cosuppression. For example, overexpression of delta-12 desaturase in soybean, rapeseed, or other oil-producing species to produce altered levels of polyunsaturated fatty acids may be achieved by expressing RNA from the full-length cDNA found in p92103, pCF2-165D, and

pSF2-169K. Transgenic lines ov rexpressing the delta-12 desaturase, when crossed with lines overexpressing delta-15 desaturases, will result in ultrahigh levels of 18:3. Similarly, the isolated nucleic acid fragments encoding fatty acid desaturases from Arabidopsis, rapeseed, and soybean can also be used by one skilled in the art to obtain other substantially homologous full-length cDNAs, if not already obtained, as well as the corresponding genes as fragments of the invention. These, in turn, may be used to overexpress the corresponding desaturases in plants. One skilled in the art can also isolate the coding sequence(s) from the fragment(s) of the invention by using and/or creating sites for restriction endonucleases, as described in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press).

One particularly useful application of the claimed inventions is to repair the agronomic performance of plant mutants containing ultra high levels of oleate in seed oil. In Arabidopsis reduction in linoleate in phosphatidylcholine due to a mutation in microsomal delta-12 desaturase affected low temperature survival [Miquel, M. et. al. (1993) Proc. Natl Acad. Sci. USA 90:6208-6212]. Furthermore, there is evidence that the poor agronomic performance of canola plants containing ultra high (>80%)levels of oleate in seed is due to mutations in the microsomal delta-12 desaturase genes that reduce the level of linoleate in phosphotidylcholine of roots and leaves. That is, the mutations are not seed-specific. Thus, the root and/or leaf-specific expression (that is, with no expression in the seeds) of microsomal delta-12 desaturase activity in mutants of oilseeds with ultra-high levels of oleate in seed oil

will result in agronomically-improved mutant plants with ultra high levels of oleate in seed oil.

Inhibition of Plant Target Genes by Use of Antisense RNA

Antisense RNA has been used to inhibit plant target genes in a tissue-specific manner (see van der Krol et al., Biotechniques (1988) 6:958-976). Antisense inhibition has been shown using the entire cDNA sequence (Sheehy et al., Proc. Natl. Acad. Sci. USA (1988) 85:8805-8809) as well as a partial cDNA sequence (Cannon et al., Plant Molec. Biol. (1990) 15:39-47). There is also evidence that the 3' non-coding sequences (Ch'ng et al., Proc. Natl. Acad. Sci. USA (1989) 86:10006-10010) and fragments of 5' coding sequence, containing as few as 41 base-pairs of a 1.87 kb cDNA (Cannon et al., Plant Molec. Biol. (1990) 15:39-47), can play important roles in antisense inhibition.

The use of antisense inhibition of the fatty acid desaturases may require isolation of the transcribed sequence for one or more target fatty acid desaturase genes that are expressed in the target tissue of the target plant. The genes that are most highly expressed are the best targets for antisense inhibition. These genes may be identified by determining their levels of transcription by techniques, such as quantitative analysis of mRNA levels or nuclear run-off transcription, known to one skilled in the art.

The entire soybean microsomal delta-12 desaturase cDNA was cloned in the antisense orientation with respect to either soybean b-conglycinin, soybean KTi3, and bean phaseolin promoter and the chimeric gene transformed into soybean somatic embryos that were previously shown to serve as good model system for soybean zygotic embryos and are predictive of seed composition (Table 11). Transformed somatic embryos

showed inhibition of linoleate biosyntheis. Similarly, the entire Brassica napus microsomal delta-12 desaturase cDNA was cloned in the antisense orientation with respect to a rapeseed napin promoter and the chimeric gene transformed into B. napus. Seeds of transformed B. napus plants showed inhibition of linoleate biosynthesis. Thus, antisense inhibition of delta-12 desaturase in oil-producing species, including corn, Brassica napus, and soybean resulting in altered levels of polyunsaturated fatty acids may be achieved by expressing antisense RNA from the entire or partial cDNA encoding microsomal delta-12 desaturase.

Inhibition of Plant Target Genes by Cosuppression

The phenomenon of cosuppression has also been used to inhibit plant target genes in a tissue-specific manner. Cosuppression of an endogenous gene using the entire cDNA sequence (Napoli et al., The Plant Cell (1990) 2:279-289; van der Krol et al., The Plant Cell (1990) 2:291-299) as well as a partial cDNA sequence (730 bp of a 1770 bp cDNA) (Smith et al., Mol. Gen. Genetics (1990) 224:477-481) are known.

The nucleic acid fragments of the instant invention encoding fatty acid desaturases, or parts thereof, with suitable regulatory sequences, can be used to reduce the level of fatty acid desaturases, thereby altering fatty acid composition, in transgenic plants which contain an endogenous gene substantially homologous to the introduced nucleic acid fragment. The experimental procedures necessary for this are similar to those described above for the overexpression of the fatty acid desaturase nucleic acid fragments except that one may also use a partial cDNA sequence. For example, cosuppression of delta-12 desaturase in Brassica napus and soybean resulting in altered levels of

polyunsaturated fatty acids may be achieved by expressing in the sense orientation the entire or partial seed delta-12 desaturase cDNA found in pCF2-165D and pSF2-165K, respectively. Endogenous genes can also be inhibited by non-coding regions of an introduced copy of the gene [For example, Brusslan, J. A. et al. (1993) Plant Cell 5:667-677; Matzke, M. A. et al., Plant Molecular Biology 16:821-830]. We have shown that an Arabidopsis gene (SEQ ID NO:15) corresponding to the cDNA (SEQ ID NO:1) can be isolated. One skilled in the art can readily isolate genomic DNA containing or flanking the genes and use the coding or non-coding regions in such transgene inhibition methods.

Analysis of the fatty acid composition of roots and seeds of Arabidopsis mutants deficient in microsomal delta-12 desaturation shows that they have reduced levels of 18:2 as well as reduced levels of 16:0 (as much as 40% reduced level in mutant seeds as compared to wild type seeds) [Miquel and Browse (1990) in Plant Lipid Biochemistry, Structure, and Utilization, pages 456-458, Ed. Quinn, P. J. and Harwood, J. L., Portland Press. Reduction in the level of 16:0 is also observed in ultra high oleate mutants of B. napus. Thus, one can expect that ultra high level of 18:1 in transgenic plants due to antisense inhibition or co-supression using the claimed sequences will also reduce the level of 16:0.

Selection of Hosts, Promoters and Enhancers

A preferred class of heterologous hosts for the expression of the nucleic acid fragments of the invention are eukaryotic hosts, particularly the cells of higher plants. Particularly preferred among the higher plants are the oil-producing species, such as soybean (Glycine max), rapeseed (including Brassica napus, B. campestris), sunflower (Helianthus annus),

cotton (Gossypium hirsutum), corn (Zea mays), cocoa (Theobroma cacao), safflower (Carthamus tinctorius), oil palm (Elaeis guineensis), coconut palm (Cocos nucifera), flax (Linum usitatissimum), and peanut (Arachis hypogaea).

Expression in plants will use regulatory sequences functional in such plants. The expression of foreign genes in plants is well-established (De Blaere et al., Meth. Enzymol. (1987) 153:277-291). The source of the promoter chosen to drive the expression of the fragments of the invention is not critical provided it has sufficient transcriptional activity to accomplish the invention by increasing or decreasing, respectively, the level of translatable mRNA for the fatty acid desaturases in the desired host tissue. Preferred promoters include (a) strong constitutive plant promoters, such as those directing the 19S and 35S transcripts in cauliflower mosaic virus (Odell et al., Nature (1985) 313:810-812; Hull et al., Virology (1987) 86:482-493), (b) tissue- or developmentally-specific promoters, and (c) other transcriptional promoter systems engineered in plants, such as those using bacteriophage T7 RNA polymerase promoter sequences to express foreign genes. Examples of tissue-specific promoters are the light-inducible promoter of the small subunit of ribulose 1,5-bis-phosphate carboxylase (if expression is desired in photosynthetic tissues), the maize zein protein promoter (Matzke et al., EMBO J. (1984) 3:1525-1532), and the chlorophyll a/b binding protein promoter (Lampa et al., Nature (1986) 316:750-752).

Particularly preferred promoters are those that allow seed-specific expression. This may be especially useful since seeds are the primary source of vegetable oils and also since seed-specific expression will avoid

any potential deleterious effect in non-seed tissues. Examples of seed-specific promoters include, but are not limited to, the promoters of seed storage proteins, which can represent up to 90% of total seed protein in many plants. The seed storage proteins are strictly regulated, being expressed almost exclusively in seeds in a highly tissue-specific and stage-specific manner (Higgins et al., Ann. Rev. Plant Physiol. (1984) 35:191-221; Goldberg et al., Cell (1989) 56:149-160). Moreover, different seed storage proteins may be expressed at different stages of seed development.

Expression of seed-specific genes has been studied in great detail (See reviews by Goldberg et al., Cell (1989) 56:149-160 and Higgins et al., Ann. Rev. Plant Physiol. (1984) 35:191-221). There are currently numerous examples of seed-specific expression of seed storage protein genes in transgenic dicotyledonous plants. These include genes from dicotyledonous plants for bean b-phaseolin (Sengupta-Gopalan et al., Proc. Natl. Acad. Sci. USA (1985) 82:3320-3324; Hoffman et al., Plant Mol. Biol. (1988) 11:717-729), bean lectin (Voelker et al., EMBO J. (1987) 6:3571-3577), soybean lectin (Okamuro et al., Proc. Natl. Acad. Sci. USA (1986) 83:8240-8244), soybean Kunitz trypsin inhibitor (Perez-Grau et al., Plant Cell (1989) 1:095-1109), soybean b-conglycinin (Beachy et al., EMBO J. (1985) 4:3047-3053; pea vicilin (Higgins et al., Plant Mol. Biol. (1988) 11:683-695), pea convicilin (Newbigin et al., Planta (1990) 180:461-470), pea legumin (Shirsat et al., Mol. Gen. Genetics (1989) 215:326-331); rapeseed napin (Radke et al., Theor. Appl. Genet. (1988) 75:685-694) as well as genes from monocotyledonous plants such as for maize 15 kD zein (Hoffman et al., EMBO J. (1987) 6:3213-3221), maize 18 kD oleosin (Lee at al., Proc. Natl. Acad. Sci. USA (1991) 888:6181-6185),

barley b-hordein (Marris et al., Plant Mol. Biol. (1988) 10:359-366) and wheat glutenin (Colot et al., EMBO J. (1987) 6:3559-3564). Moreover, promoters of seed-specific genes operably linked to heterologous coding sequences in chimeric gene constructs also maintain their temporal and spatial expression pattern in transgenic plants. Such examples include use of Arabidopsis thaliana 2S seed storage protein gene promoter to express enkephalin peptides in Arabidopsis and B. napus seeds (Vandekerckhove et al., Bio/Technology (1989) 7:929-932), bean lectin and bean b-phaseolin promoters to express luciferase (Riggs et al., Plant Sci. (1989) 63:47-57), and wheat glutenin promoters to express chloramphenicol acetyl transferase (Colot et al., EMBO J. (1987) 6:3559-3564).

Of particular use in the expression of the nucleic acid fragment of the invention will be the heterologous promoters from several soybean seed storage protein genes such as those for the Kunitz trypsin inhibitor (Jofuku et al., Plant Cell (1989) 1:1079-1093; glycinin (Nielson et al., Plant Cell (1989) 1:313-328), and b-conglycinin (Harada et al., Plant Cell (1989) 1:415-425). Promoters of genes for a- and b-subunits of soybean b-conglycinin storage protein will be particularly useful in expressing the mRNA or the antisense RNA in the cotyledons at mid- to late-stages of seed development (Beachy et al., EMBO J. (1985) 4:3047-3053) in transgenic plants. This is because there is very little position effect on their expression in transgenic seeds, and the two promoters show different temporal regulation. The promoter for the a-subunit gene is expressed a few days before that for the b-subunit gene. This is important for transforming rapeseed where oil biosynthesis begins about a week

before seed storage protein synthesis (Murphy et al., J. Plant Physiol. (1989) 135:63-69).

Also of particular use will be promoters of genes expressed during early embryogenesis and oil biosynthesis. The native regulatory sequences, including the native promoters, of the fatty acid desaturase genes expressing the nucleic acid fragments of the invention can be used following their isolation by those skilled in the art. Heterologous promoters from other genes involved in seed oil biosynthesis, such as those for B. napus isocitrate lyase and malate synthase (Comai et al., Plant Cell (1989) 1:293-300), delta-9 desaturase from safflower (Thompson et al. Proc. Natl. Acad. Sci. USA (1991) 88:2578-2582) and castor (Shanklin et al., Proc. Natl. Acad. Sci. USA (1991) 88:2510-2514), acyl carrier protein (ACP) from Arabidopsis (Post-Beittenmiller et al., Nucl. Acids Res. (1989) 17:1777), B. napus (Safford et al., Eur. J. Biochem. (1988) 174:287-295), and B. campestris (Rose et al., Nucl. Acids Res. (1987) 15:7197), b-ketoacyl-ACP synthetase from barley (Siggaard-Andersen et al., Proc. Natl. Acad. Sci. USA (1991) 88:4114-4118), and oleosin from Zea mays (Lee et al., Proc. Natl. Acad. Sci. USA (1991) 88:6181-6185), soybean (Genbank Accession No: X60773) and B. napus (Lee et al., Plant Physiol. (1991) 96:1395-1397) will be of use. If the sequence of the corresponding genes is not disclosed or their promoter region is not identified, one skilled in the art can use the published sequence to isolate the corresponding gene and a fragment thereof containing the promoter. The partial protein sequences for the relatively-abundant enoyl-ACP reductase and acetyl-CoA carboxylase are also published (Slabas et al., Biochim. Biophys. Acta (1987) 877:271-280; Cottingham et al., Biochim. Biophys. Acta (1988) 954:201-207) and one skilled in the art can use

these sequences to isolate the corresponding seed genes with their promoters. Similarly, the fragments of the present invention encoding fatty acid desaturases can be used to obtain promoter regions of the corresponding genes for use in expressing chimeric genes.

Attaining the proper level of expression of the nucleic acid fragments of the invention may require the use of different chimeric genes utilizing different promoters. Such chimeric genes can be transferred into host plants either together in a single expression vector or sequentially using more than one vector.

It is envisioned that the introduction of enhancers or enhancer-like elements into the promoter regions of either the native or chimeric nucleic acid fragments of the invention will result in increased expression to accomplish the invention. This would include viral enhancers such as that found in the 35S promoter (Odell et al., Plant Mol. Biol. (1988) 10:263-272), enhancers from the opine genes (Fromm et al., Plant Cell (1989) 1:977-984), or enhancers from any other source that result in increased transcription when placed into a promoter operably linked to the nucleic acid fragment of the invention.

Of particular importance is the DNA sequence element isolated from the gene for the a-subunit of b-conglycinin that can confer 40-fold seed-specific enhancement to a constitutive promoter (Chen et al., Dev. Genet. (1989) 10:112-122). One skilled in the art can readily isolate this element and insert it within the promoter region of any gene in order to obtain seed-specific enhanced expression with the promoter in transgenic plants. Insertion of such an element in any seed-specific gene that is expressed at different times than the b-conglycinin gene will result in expression in

transgenic plants for a longer period during seed development.

The invention can also be accomplished by a variety of other methods to obtain the desired end. In one form, the invention is based on modifying plants to produce increased levels of fatty acid desaturases by virtue of introducing more than one copy of the foreign gene containing the nucleic acid fragments of the invention. In some cases, the desired level of polyunsaturated fatty acids may require introduction of foreign genes for more than one kind of fatty acid desaturase.

Any 3' non-coding region capable of providing a polyadenylation signal and other regulatory sequences that may be required for the proper expression of the nucleic acid fragments of the invention can be used to accomplish the invention. This would include 3' ends of the native fatty acid desaturase(s), viral genes such as from the 35S or the 19S cauliflower mosaic virus transcripts, from the opine synthesis genes, ribulose 1,5-bisphosphate carboxylase, or chlorophyll a/b binding protein. There are numerous examples in the art that teach the usefulness of different 3' non-coding regions.

Transformation Methods

Various methods of transforming cells of higher plants according to the present invention are available to those skilled in the art (see EPO Pub. 0 295 959 A2 and 0 318 341 A1). Such methods include those based on transformation vectors utilizing the Ti and Ri plasmids of Agrobacterium spp. It is particularly preferred to use the binary type of these vectors. Ti-derived vectors transform a wide variety of higher plants, including monocotyledonous and dicotyledonous plants (Sukhapinda et al., Plant Mol. Biol. (1987) 8:209-216; Potrykus, Mol. Gen. Genet. (1985) 199:183). Other

transformation methods are available to those skilled in the art, such as direct uptake of foreign DNA constructs (see EPO Pub. 0 295 959 A2), techniques of electroporation (Fromm et al., Nature (1986) (London) 319:791) or high-velocity ballistic bombardment with metal particles coated with the nucleic acid constructs (Kline et al., Nature (1987) (London) 327:70). Once transformed, the cells can be regenerated by those skilled in the art.

Of particular relevance are the recently described methods to transform foreign genes into commercially important crops, such as rapeseed (De Block et al., Plant Physiol. (1989) 91:694-701), sunflower (Everett et al., Bio/Technology (1987) 5:1201), and soybean (Christou et al., Proc. Natl. Acad. Sci USA (1989) 86:7500-7504.

Application to Molecular Breeding

The 1.6 kb insert obtained from the plasmid pSF2-169K was used as a radiolabelled probe on a Southern blot containing genomic DNA from soybean (*Glycine max* (cultivar Bonus) and *Glycine soja* (PI81762)) digested with one of several restriction enzymes. Different patterns of hybridization (polymorphisms) were identified in digests performed with restriction enzymes Hind III and Eco RI. These polymorphisms were used to map two pSF2-169 loci relative to other loci on the soybean genome essentially as described by Helentjaris et al., (Theor. Appl. Genet. (1986) 72:761-769). One mapped to linkage group 11 between 4404.00 and 1503.00 loci (4.5 cM and 7.1 cM from 4404.00 and 1503.00, respectively) and the other to linkage group 19 between 4010.00 and 5302.00 loci (1.9 cM and 2.7 cM from 4010.00 and 5302.00, respectively) [Rafalski, A and Tingey, S. (1993) in Genetic Maps, Ed. O' Brien, S. J.]. The use of

restriction fragment length polymorphism (RFLP) markers in plant breeding has been well-documented in the art (Tanksley et al., Bio/Technology (1989) 7:257-264). Thus, the nucleic acid fragments of the invention can be used as RFLP markers for traits linked to expression of fatty acid desaturases. These traits will include altered levels of unsaturated fatty acids. The nucleic acid fragment of the invention can also be used to isolate the fatty acid desaturase gene from variant (including mutant) plants with altered levels of unsaturated fatty acids. Sequencing of these genes will reveal nucleotide differences from the normal gene that cause the variation. Short oligonucleotides designed around these differences may also be used in molecular breeding either as hybridization probes or in DNA-based diagnostics to follow the variation in fatty acids. Oligonucleotides based on differences that are linked to the variation may be used as molecular markers in breeding these variant oil traits.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. All publications, including patents and non-patent literature, referred to in this specification are expressly incorporated by reference herein.

EXAMPLE 1

ISOLATION OF GENOMIC DNA FLANKING THE T-DNA SITE OF INSERTION IN ARABIDOPSIS THALIANA MUTANT LINE 658

Identification of an Arabidopsis thaliana T-DNA Mutant with High Oleic Acid Content

A population of Arabidopsis thaliana (geographic race Wassilewskija) transformants containing the modified T-DNA of Agrobacterium tumefaciens was generated by seed transformation as described by Feldmann et al., (Mol. Gen. Genetics (1987) 208:1-9). In this population the transformants contain DNA sequences encoding the pBR322 bacterial vector, nopaline synthase, neomycin phosphotransferase (NPTII, confers kanamycin resistance), and b-lactamase (confers ampicillin resistance) within the T-DNA border sequences. The integration of the T-DNA into different areas of the chromosomes of individual transformants may cause a disruption of plant gene function at or near the site of insertion, and phenotypes associated with this loss of gene function can be analyzed by screening the population for the phenotype.

T3 seed was generated from the wild type seed treated with Agrobacterium tumefaciens by two rounds of self-fertilization as described by Feldmann et al., (Science (1989) 243:1351-1354). These progeny were segregating for the T-DNA insertion, and thus for any mutation resulting from the insertion. Approximately 10-12 leaves of each of 1700 lines were combined and the fatty acid content of each of the 1700 pooled samples was determined by gas chromatography of the fatty acyl methyl esters essentially as described by Browse et al., (Anal. Biochem. (1986) 152:141-145) except that 2.5% H_2SO_4 in methanol was used as the methylation reagent. A line designated "658" produced a sample that gave an

alt red fatty acid profile compared to those of lines 657 and 659 sampled at the same time (Table 8).

TABLE 8

Fatty Acid Methyl Ester	657 Leaf Pool	659 Leaf Pool	658 Leaf Pool
16:0	14.4	14.1	13.6
16:1	4.4	4.6	4.5
16:2	2.9	2.2	2.7
16:3	13.9	13.3	13.9
18:0	1.0	1.1	0.9
18:1	2.6	2.5	4.9
18:2	14.0	13.6	12.8
18:3	42.9	46.1	44.4

Analysis of the fatty acid composition of 12 individual T3 seeds of line 658 indicated that the 658 pool was composed of seeds segregating in three classes: "high", "mid-range" and "low" classes with approximately 37% (12 seeds), 21% (7 seeds), and 14% (3 seeds) oleic acid, respectively (Table 9).

TABLE 9

	"High" Class	"Mid-range" Class	"Low" Class
16:0	8.9	8.7	9.3
16:1c	2.0	1.6	2.6
18:0	4.5	4.3	4.4
18:1	37.0	20.7	14.4
18:2	8.0	24.9	27.7
18:3	10.6	14.3	13.6
20:1	25.5	21.6	20.4

Thus, the high oleic acid mutant phenotype segregates in an approximately Mendelian ratio. To determine the number of independently segregating T-DNA

inserts in line 658, 200 T3 seeds were tested for their ability to germinate and grow in the presence of kanamycin [Feldman et al. (1989) Science 243:1351-1354]. In this experiment, only 4 kanamycin-sensitive individual plants were identified. The segregation ratio (approximately 50:1) indicated that line 658 harbored three T-DNA inserts. In this and two other experiments a total of 56 kanamycin-sensitive plants were identified; 53 of these were analyzed for fatty acid composition and at least seven of these displayed oleic acid levels that were higher than would be expected for wild type seedlings grown under these conditions.

In order to more rigorously test whether the mutation resulting in high oleic acid is the result of T-DNA insertion, Applicants identified a derivative line that was segregating for the mutant fatty acid phenotype as well as a single kanamycin resistance locus. For this, approximately 100 T3 plants were individually grown to maturity and seeds collected. One sample of seed from each T3 plant was tested for the ability to germinate and grow in the presence of kanamycin. In addition, the fatty acid compositions of ten additional individual seeds from each line were determined. A T3 plant, designated 658-75, was identified whose progeny seeds segregated 28 kanamycin-sensitive to 60 kanamycin-resistant and 7 with either low or intermediate oleic acid to 2 high oleic acid.

A total of approximately 400 T4 progeny seeds of the derivative line 658-75 were grown and the leaf fatty acid composition analyzed. A total of 91 plants were identified as being homozygous for the high oleic acid trait (18:2/18:1 less than 0.5). The remaining plants (18:2/18:1 more than 1.2) could not be definitively assigned to wild type and heterozygous classes on the

basis of leaf fatty acid composition and thus could not be used to test linkage between the kanamycin marker and the fatty acid mutation. Eighty three of the 91 apparently homozygous high oleic acid mutant were tested for the presence of nopaline, another T-DNA marker, in leaf extracts (Errampalli et al. The Plant Cell (1991)3:149-157 and all 83 plants were positive for the presence of nopaline. This tight linkage of the mutant fatty acid phenotype and a T-DNA marker provides evidence that the high oleic acid trait in mutant 658 is the result of T-DNA insertion.

Plasmid Rescue and Analysis

One-half and one microgram of genomic DNA from the homozygous mutant plants of the 658-75 line, prepared from leaf tissue as described [Rogers, S. O. and A. J. Bendich (1985) Plant Molecular Biology 5:69-76], was digested with 20 units of either Bam HI or Sal I restriction enzyme (Bethesda Research Laboratory) in a 50 µL reaction volume according to the manufacturer's specifications. After digestion the DNA was extracted with buffer-saturated phenol (Bethesda Research Laboratory) followed by precipitation in ethanol. One-half to one microgram of Bam HI or Sal I digested genomic DNA was resuspended in 200 uL or 400 uL of ligation buffer containing 50 mM Tris-Cl, pH 8.0, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP, and 4 units of T4 DNA ligase (Bethesda Research Laboratory). The dilute DNA concentration of approximate 2.5 ug/mL in the ligation reaction was chosen to facilitate circularization, as opposed to intermolecular joining. The reaction was incubated for 16 h at 16°C. Competent DH10B cells (Bethesda Research Laboratory) were transfected with 10 ng of ligated DNA per 100 µL of competent cells according to the manufacturer's specifications. Transformants from Sal I or Bam HI

digests were selected on LB plates (10 g Bacto-tryptone, 5 g Bacto-yeast extract, 5 g NaCl, 15 g agar per liter, pH 7.4) containing 100 µg/mL ampicillin. After overnight incubation at 37°C the plates were scored for ampicillin-resistant colonies.

A single ampicillin-resistant transformant derived from Bam HI-digested plant DNA was used to start a culture in 35 mL LB medium (10 g Bacto-tryptone, 5 g yeast-extract, 5 g NaCl per liter) containing 25 mg/L ampicillin. The culture was incubated with shaking overnight at 37°C and the cells were then collected by centrifugation at 1000xg for 10 min. Plasmid DNA, designated p658-1, was isolated from the cells by the alkaline lysis method of Birmbiom et al. [Nucleic Acid Research (1979) 7:1513-1523], as described in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed (1989) Cold Spring Harbor Laboratory Press). Plasmid p658-1 DNA was digested by restriction enzymes Bam HI, Eco RI and Sal I (Bethesda Reseach Laboratory) and electrophoresed through a 1% agarose gel in 1xTBE buffer (0.089M tris-borate, 0.002M EDTA). The restriction pattern indicated the presence in this plasmid of the expected 14.2 kB T-DNA fragment and a 1.6 kB putative plant DNA/T-DNA border fragment.

EXAMPLE 2

CLONING OF ARABIDOPSIS THALIANA MICROSOMAL DELTA-12 DESATURASE cDNA USING GENOMIC DNA FLANKING THE T-DNA SITE OF INSERTION IN ARABIDOPSIS THALIANA MUTANT LINE 658-75 AS A HYBRIDIZATION PROBE

Two hundred nanograms of the 1.6 kB Eco RI-Bam HI fragment from plasmid p658-1, following digestion of the plasmid with Eco RI and Bam HI and purification by electrophoresis in agarose, was radiolabelled with alpha[32P]-dCTP using a Random Priming Labeling Kit

(Bethesda Research Laboratory) under conditions recommended by the manufacturer.

The radiolabeled DNA was used as a probe to screen an Arabidopsis cDNA library made from RNA isolated from above ground portions of various growth stages (Elledge et al., (1991) Proc. Nat. Acad. Sci., 88:1731-1735) essentially as described in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press). For this, approximately 17,000 plaque-forming units were plated on seven 90mm petri plates containing a lawn of LE392 E. coli cells on NZY agar media (5 g NaCl, 2 g MgSO4-7 H20, 5 g yeast extract, 10 g casein acid hydrolysate, 13 g agar per liter). Replica filters of the phage plaques were prepared by adsorbing the plaques onto nitrocellulose filters (BA85, Schleicher and Schuell) then soaking successively for five min each in 0.5 M NaOH/1 M NaCl, 0.5 M Tris(pH 7.4)/1.5 M NaCl and 2xSSPE (0.36 M NaCl, 20 mM NaH2PO4(p H7.4), 20 mM EDTA (pH 7.4)). The filters were then air dried and baked for 2 h at 80°C. After baking the filters were wetted in 2X SSPE, and then incubated at 42°C in prehybridization buffer (50% Formamide, 5X SSPE, 1% SDS, 5X Denhardt's Reagent, and 100 ug/mL denatured salmon sperm DNA) for 2 h. The filters were removed from the prehybridization buffer, and then transferred to hybridization buffer (50% Formamide, 5X SSPE, 1% SDS, 1X Denhardt's Reagent, and 100 ug/mL denatured salmon sperm DNA) containing the denatured radiolabeled probe (see above) and incubated for 40 h at 42°C. The filters were washed three times in 2X SSPE/0.2% SDS at 42°C (15 min each) and twice in 0.2X SSPE/0,2% SDS at 55°C (30 min each), followed by autoradiography on Kodak XAR-5 film with an intensifying screen at -80°C, overnight. Fifteen plaques were identified as positively-

hybridizing on replica filters. Five of these were subjected to plaque purification essentially as described in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd ed. (1989), Cold Spring Harbor Laboratory Press). The lambda YES-R cDNA clones were converted to plasmid by propagating the phage in the E. coli BNN--132 cells, which expresses Cre protein that excises the cDNA insert as a double-stranded plasmid by cre-mediated in vivo site-speicifc recombination at a 'lox' sites present in the phage. Ampicillin-resistant plasmid clones containing cDNA inserts were grown in liquid culture, and plasmid DNA was prepared using the alkaline lysis method as previously described. The sizes of the resulting plasmids were analyzed by electrophoresis in agarose gels. The agarose gels were treated with 0.5 M NaOH/1 M NaCl, and 0.5 M Tris(pH 7.4), 1.5 M NaCl for 15 min each, and the gel was then dried completely on a gel drier at 65°C. The gel was hydrated in 2X SSPE and incubated overnight, at 42°C, in hybridization buffer containing the denatured radiolabeled probe, followed by washing as described above. After autoradiography, the inserts of four of the purified cDNA clones were found to have hybridized to the probe. Plasmid DNA from the hybridizing clones was purified by equilibration in a CsCl/ethidium bromide gradient (see above). The four cDNA clones were sequenced using Sequenase T7 DNA polymerase (US Biochemical Corp.) and the manufacturer's instructions, beginning with primers homologous to vector sequences that flank the cDNA insert. After comparing the partial sequences of the inserts obtained from the four clones, it was apparent that they each contained sequences in common. One cDNA clone, p92103, containing ca. 1.4 kB cDNA insert, was sequenced. The longest three clones were subcloned into the plasmid vector pBluescript

(Stratagene). One of these clones, designated pSF2b, containing ca 1.2 kB cDNA insert was also sequenced serially with primers designed from the newly acquired sequences as the sequencing experiment progressed. The composite sequence derived from pSF2b and p92103 is shown in SEQ ID NO:1.

EXAMPLE 3

CLONING OF PLANT FATTY ACID DESATURASE cDNAs USING THE ARABIDOPSIS THALIANA MICROSOMAL DELTA-12 DESATURASE cDNA CLONE AS A HYBRIDIZATION PROBE

An approximately 1.2 kb fragment containing the *Arabidopsis* delta-12 desaturase coding sequence of SEQ ID NO:1 was obtained from plasmid pSF2b. This plasmid was digested with EcoR I and the 1.2 kb delta-12 desaturase cDNA fragment was purified from the vector sequence by agarose gel electrophoresis. The fragment was radiolabelled with ^{32}P as previously described.

Cloning of a Brassica napus Seed cDNA Encoding Microsomal Delta-12 Fatty Acid Desaturase

The radiolabelled probe was used to screen a *Brassica napus* seed cDNA library. In order to construct the library, *Brassica napus* seeds were harvested 20-21 days after pollination, placed in liquid nitrogen, and polysomal RNA was isolated following the procedure of Kamalay et al., (Cell (1980) 19:935-946). The polyadenylated mRNA fraction was obtained by affinity chromatography on oligo-dT cellulose (Aviv et al., Proc. Natl. Acad. Sci. USA (1972) 69:1408-1411). Four micrograms of this mRNA were used to construct a seed cDNA library in lambda phage (Uni-ZAP™ XR vector) using the protocol described in the ZAP-cDNA™ Synthesis Kit (1991 Stratagene Catalog, Item #200400). Approximately 600,000 clones were screened for positively hybridizing

plaques using the radiolabelled EcoR I fragm nt from pSF2b as a probe essentially as described in Sambrook et al., (Molecular Cloning: A Laboratory Manual, 2nd ed. (1989) Cold Spring Harbor Laboratory Press) except that low stringency hybridization conditions (50 mM Tris, pH 7.6, 6X SSC, 5X Denhardt's, 0.5% SDS, 100 µg denatured calf thymus DNA and 50°C) were used and post-hybridization washes were performed twice with 2X SSC, 0.5% SDS at room temperature for 15 min, then twice with 0.2X SSC, 0.5% SDS at room temperature for 15 min, and then twice with 0.2X SSC, 0.5% SDS at 50°C for 15 min. Ten positive plaques showing strong hybridization were picked, plated out, and the screening procedure was repeated. From the secondary screen nine pure phage plaques were isolated. Plasmid clones containing the cDNA inserts were obtained through the use of a helper phage according to the in vivo excision protocol provided by Stratagene. Double-stranded DNA was prepared using the alkaline lysis method as previously described, and the resulting plasmids were size-analyzed by electrophoresis in agarose gels. The largest one of the nine clones, designated pCF2-165D, contained an approximately 1.5 kb insert which was sequenced as described above. The sequence of 1394 bases of the cDNA insert of pCF2-165D is shown in SEQ ID NO:3. Contained in the insert but not shown in SEG ID NO:3 are approximately 40 bases of the extreme 5' end of the 5' non-translated region and a poly A tail of about 38 bases at the extreme 3' end of the insert.

Cloning of a Soybean Seed cDNA Encoding Microsomal Delta-12 Fatty Acid Desaturase

A cDNA library was made as follows: Soybean embryos (ca. 50 mg fresh weight each) were removed from the pods and frozen in liquid nitrogen. The frozen

embryos were ground to a fine powder in the presence of liquid nitrogen and then extracted by Polytron homogenization and fractionated to enrich for total RNA by the method of Chirgwin et al. (Biochemistry (1979) 18:5294-5299). The nucleic acid fraction was enriched for poly A$^+$RNA by passing total RNA through an oligo-dT cellulose column and eluting the poly A$^+$RNA with salt as described by Goodman et al. (Meth. Enzymol. (1979) 68:75-90). cDNA was synthesized from the purified poly A$^+$RNA using cDNA Synthesis System (Bethesda Research Laboratory) and the manufacturer's instructions. The resultant double-stranded DNA was methylated by Eco RI DNA methylase (Promega) prior to filling-in its ends with T4 DNA polymerase (Bethesda Research Laboratory) and blunt-end ligation to phosphorylated Eco RI linkers using T4 DNA ligase (Pharmacia). The double-stranded DNA was digested with Eco RI enzyme, separated from excess linkers by passage through a gel filtration column (Sepharose CL-4B), and ligated to lambda ZAP vector (Stratagene) according to manufacturer's instructions. Ligated DNA was packaged into phage using the Gigapack packaging extract (Stratagene) according to manufacturer's instructions. The resultant cDNA library was amplified as per Stratagene's instructions and stored at -80°C.

Following the instructions in the Lambda ZAP Cloning Kit Manual (Stratagene), the cDNA phage library was used to infect _E. coli_ BB4 cells and approximately 600,000 plaque forming units were plated onto 150 mm diameter petri plates. Duplicate lifts of the plates were made onto nitrocellulose filters (Schleicher & Schuell). The filters were prehybridized in 25 mL of hybridization buffer consisting of 6X SSPE, 5X Denhardt's solution, 0.5% SDS, 5% dextran sulfate and 0.1 mg/mL denatured salmon sperm DNA (Sigma Chemical

Co.) at 50°C for 2 h. Radiolabelled probe prepared from pSF2b as described above was added, and allowed to hybridize for 18 h at 50°C. The filters were washed exactly as described above. Autoradiography of the filters indicated that there were 14 strongly hybridizing plaques. The 14 plaques were subjected to a second round of screening as before. Numerous, strongly hybridizing plaques were observed on 6 of the 14 filters, and one, well-isolated from other phage, was picked from each of the six plates for further analysis.

Following the Lambda ZAP Cloning Kit Instruction Manual (Stratagene), sequences of the pBluescript vector, including the cDNA inserts, from the purified phages were excised in the presence of a helper phage and the resultant phagemids were used to infect E. coli XL-1 Blue cells. DNA from the plasmids was made by the Promega "Magic Miniprep" according to the manufacturers instructions. Restriction analysis indicated that the plasmids contained inserts ranging in size from 1 kb to 2.5 kb. The alkali-denatured double-stranded DNA from one of these, designated pSF2-169K contained an insert of 1.6 kb, was sequenced as described above. The nucleotide sequence of the cDNA insert in plasmid pSF2-169K shown in SEQ ID NO:5.

Cloning of a Corn (Zea mays)
cDNA Encoding Seed Microsomal Delta-12
Fatty Acid Desaturase

Corn microsomal delta-12 desaturase cDNA was isolated using a PCR approach. For this, a cDNA library was made to poly A^+ RNA from developing corn embryos in Lambda ZAP II vector (Stratagene). 5-10 ul of this library was used as a template for PCR using 100 pmol each of two sets of degenerate oligomers NS3 (SEQ ID NO:13) and equimolar amounts of RB5a/b (that is, equimolar amounts of SEQ ID NOS:16/17) as sense and

antisense primers, respectively. NS3 and RB5a/b correspond to stretches of amino acids 101-109 and 318-326, respectively, of SEQ ID NO:2, which are conserved in most microsomal delta-12 desaturases (SEQ ID NOS:2, 4, 6, 8). PCR was carried out using the PCR kit (Perkin-Elmer) using 40 cycles of 94°C 1 min, 45°C, 1 min, and 55°C, 2 min. Analyses of the PCR products on an agarose gel showed the presence of a product of the expected size (720 bp), which was absent in control reactions containing either the sense or antisense primers alone. The PCR product fragment was gel purified and then used as a probe for screening the same corn cDNA library at 60°C as described above. One positively-hybridizing plaque was purified and partial sequence determination of its cDNA showed it to be a nucleotide sequence encoding microsomal delta-12 desaturase but truncated at the 3' end. The cDNA insert encoding the partial desaturase was gel isolated and used to probe the corn cDNA library again. Several positive plaques were recovered and characterized. DNA sequence analysis revealed that all of these clones seem to represent the same sequence with the different length of 5' or 3' ends. The clone containing the longest insert, designated pFad2#1, was sequenced completely. SEQ ID NO:7 shows the 5' to 3' nucleotide sequence of 1790 base pairs of corn (Zea mays) cDNA which encodes microsomal delta-12 desaturase in plasmid pFad2#1. Nucleotides 165 to 167 and nucleotides 1326 to 1328 are, respectively, the putative initiation codon and the termination codon of the open reading frame (nucleotides 164 to 1328). SEQ ID NO:8 is the 387 amino acid protein sequence deduced from the open reading frame (nucleotides 164 to 1328) in SEQ ID NO:7. The deduced amino acid sequence of the polypeptide shared overall identities of 71%, 40%, and 38% to Arabidopsis

microsomal delta-12 desaturase, Arabidopsis microsomal delta-15 desaturase, and Arabidopsis plastid delta-15 desaturase, respectively. Furthermore, it lacked an N-terminal amino acid extension that would indicate it is a plastid enzyme. Based on these considerations, it is concluded that it encodes a microsomal delta-12 desaturase.

Cloning of a cDNA Encoding A Microsomal Delta-12 Desaturase and of cDNAs Encoding Microsomal Delta-12 Desaturase-Related Enzymes from Castor Bean Seed

Castor microsomal delta-12 desaturase cDNA was isolated using a RT-PCR approach. Polysomal mRNA was isolated from castor beans of stages I-II (5-10 DAP) and also from castor beans of stages IV-V (20-25 DAP). Ten ng of each mRNA was used for separate RT-PCR reactions, using the Perkin-Elmer RT-PCR kit with the reagent concentration as recommended by the kit protocol. The reverse transcriptase reaction was primed with random hexamers and the PCR reaction with 100 pmol each of the degenerate delta-12 desaturase primers NS3 and NS9 (SEQ ID NOS:13 and 14, respectively). The reverse transcriptase reaction was incubated at 25°C for 10 min, 42°C for 15 min, 99°C for 5 min and 5°C for 5 min. The PCR reaction was incubated at 95°C for 2 min followed by 35 cycles of 95°C for 1 min/50°C for 1 min. A final incubation at 60°C for 7 min completed the reaction. A DNA fragment of 720 bp was amplified from both stage I-II and stage IV-V mRNA. The amplified DNA fragment from one of the reactions was gel purified and cloned into a pGEM-T vector using the Promega pGEM-T PCR cloning kit to create the plasmid pRF2-1C. The 720 bp insert in pRF2-1C was sequenced, as described above, and the resulting DNA sequence is shown in SEQ ID NO:9. The DNA sequence in SEQ ID NO:9 contains an open-reading frame encoding 219 amino acids (SEQ ID NO:10), which has

81% identity (90% similarity) with amino acids 135 to 353 of the _Arabidopsis_ microsomal delta-12 desaturase described in SEQ ID NO:2. The cDNA insert in pRF2-1C is therefore a 673 bp fragment of a full-length cDNA encoding a castor bean seed microsomal delta-12 desaturase. The full length castor bean seed microsomal delta-12 desaturase cDNA may isolated by screening a castor seed cDNA library, at 60°C, with the labeled insert of pRF2-1C as described in the example above. The insert in pRF2-1C may also be used to screen castor bean libraries at lower temperatures to isolate delta-12 desaturase related sequences, such as the delta-12 hydroxylase.

A cDNA library made to poly A^+ mRNA isolated from developing castor beans (stages IV-V, 20-25 DAP) was screened as described above. Radiolabeled probe prepared from pSF2b or pRF2-1C, as described above, were added, and allowed to hybridize for 18 h at 50°C. The filters were washed as described above. Autoradiography of the filters indicated that there were numerous hybridizing plaques, which appeared either strongly hybridising or weakly hybridising. Three of the strongly hybridisng plaques (190A-41, 190A-42 and 190A-44) and three of the weakly hybridising plaques, (190B-41, 190b-43 and 197c-42), were plaque purified using the methods described above. The cDNA insert size of the purified phages were determined by PCR amplication of the insert using phage as template and lambda-gt11 oligomers (Clontech lambda-gt11 Amplimers) for primers. The PCR-amplified inserts of the amplified phages were subcloned into pBluescript (Pharmacia) which had been cut with Eco RI and filled in with Klenow (Sambrook et al. (Molecular Cloning, A Laboratory Approach, 2nd. ed. (1989) Cold Spring Harbor Laboratory Press). The resulting plasmids were called pRF190a-41,

pRF190a-42, pRF190a-44, pRF190b-41, pRF190b-43 and pRF197c-42. All of the inserts were about 1.1 kb with the exception of pRF197c-42 which was approx. 1.5 kb. The inserts in the plasmids were sequenced as described above. The insert in pRF190b-43 did not contain any open reading frame and was not identified. The inserts in pRF190a-41, pRF190a-42, pRF190a-44 and pRF190b-41 were identical. The insert in pRF197c-42 contained all of the nucleotides of the inserts in pRF190a-41, pRF190a-42, pRF190a-44 and pRF190b-41 plus an additional approx. 400 bp. It was deduced therefore that the insert in pRF197c-42 was a longer version of the inserts in pRF190a-41, pRF190a-42, pRF190a-44 and pRF190b-41 and all were derived from the same full-length mRNA. The complete cDNA sequence of the insert in plasmid pRF197c-42 is shown in SEQ ID NO:11. The deduced amino acid sequence of SEQ ID NO:11, shown in SEQ ID NO:12, is 78.5% identical (90% similarity) to the castor microsomal delta-12 desaturase described above (SEQ ID NO:10) and 66% identical (80% similarity) to the Arabidopsis delta-12 desaturase amino acid sequence in SEQ ID NO:2. These similarities confirm that pRF197c-42 is a castor bean seed cDNA that encodes a microsomal delta-12 desaturase or a microsomal delta-12 desaturase-related enzyme, such as a delta-12 hydroxylase. Specific PCR primers for pRF2-1C and pRF197c-42 were made. For pRF2-1c the upstream primer was bases 180 to 197 of the cDNA sequence in SEQ ID NO:9. For pRF197c-42 the upstream primer was bases 717 to 743 of the cDNA sequence in SEQ ID NO:11. A common downstream primer was made corresponding to the exact complement of the nucleotides 463 to 478 of the sequence described in SEQ ID NO:9. Using RT-PCR with random hexamers and the above primers, and the incubation temperatures described above, it was observed that mRNA which gave rise to the

cDNA contained in pRF2-1C is present in both Stage I-II and Stage IV-V castor bean seeds whereas mRNA which gave rise to the cDNA contained in plasmid pRF197c-42 is present only in Stage IV-V castor bean seeds, i.e., it is only expressed in tissue actively synthesizing ricinoleic acid. Thus it is possible that this cDNA encodes a delta-12 hydroxylase.

Clones such as pRF2-1C and pRF197c-42, and other clones from the differential screening, which, based on their DNA sequence, are less related to castor bean seed microsomal delta-12 desaturases and are not any of the known fatty-acid desaturases described above or in WO 9311245, may be expressed, for example, in soybean embryos or another suitable plant tissue, or in a microorganism, such as yeast, which does not normally contain ricinoleic acid, using suitable expression vectors and transformation protocols. The presence of novel ricinoleic acid in the transformed tissue(s) expressing the castor cDNA would confirm the identity of the castor cDNA as DNA encoding for an oleate hydroxylase.

EXAMPLE 4

USE OF THE ARABIDOPSIS THALIANA DELTA-12 DESATURASE GENOMIC CLONE AS A RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP) MARKER TO MAP THE DELTA-12 DESATURASE LOCUS IN ARABIDOPSIS

The gene encoding *Arabidopsis* microsomal delta-12 desaturase was used to map the genetic locus encoding the microsomal delta-12 desaturase of *Arabidopsis thaliana*. pSF2b cDNA insert encoding *Arabidopsis* microsomal delta-12 desaturase DNA was radiolabeled and used to screen an *Arabidopsis* genomic DNA library. DNA from several pure strongly-hybridizing phages was isolated. Southern blot analysis of the DNA from different phages using radiolabeled pSF2b cDNA insert as

the probe identified a 6 kb Hind III insert fragment to contain the coding region of the gene. This fragment was subcloned in pBluescript vector to result in plasmid pAGF2-6 and used for partial sequence determination. This sequence (SEQ ID NO:15) confirmed that it is the microsomal delta-12 desaturase gene. DNA from two phages was isolated and labelled with ^{32}P using a random priming kit from Pharmacia under conditions recommended by the manufacturer. The radioactive DNA was used to probe a Southern blot containing genomic DNA from Arabidopsis thaliana (ecotype Wassileskija and marker line W100 ecotype Landesberg background) digested with one of several restriction endonucleases. Following hybridization and washes under standard conditions (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed. (1989) Cold Spring Harbor Laboratory Press), autoradiograms were obtained. A different pattern of hybridization (polymorphism) was identified in Hind III-digested genomic DNAs using one of the phage DNAs. This polymorphism was located to a 7 kB Hind III fragment in the phage DNA that revealed the polymorphism. The 7 kb fragment was subcloned in pBluescript vector to result in plasmid pAGF2-7. Plasmid pAGF2-7 was restricted with Hind III enzyme and used as a radiolabelled probe to map the polymorphism essentially as described by Helentjaris et al., (Theor. Appl. Genet. (1986) 72:761-769). The radiolabelled DNA fragment was applied as described above to Southern blots of Hind III-digested genomic DNA isolated from 117 recombinant inbred progeny (derived from single-seed descent lines to the F_6 generation) resulting from a cross between Arabidopsis thaliana marker line W100 and ecotype Wassileskija (Burr et al., Genetics (1988) 118:519-526). The bands on the autoradiograms were interpreted as resulting from inheritance of either

pat rnal (ecotype Wassileskija) or maternal (marker line W100) DNA or both (a heterozygote). The resulting segregation data were subjected to genetic analysis using the computer program Mapmaker (Lander et al., Genomics (1987) 1:174-181). In conjunction with previously obtained segregation data for 63 anonymous RFLP markers and 9 morphological markers in Arabidopsis thaliana (Chang et al., Proc. Natl. Acad. Sci. USA (1988) 85:6856-6860; Nam et al., Plant Cell (1989) 1:699-705), a single genetic locus was positioned corresponding to the microsomal delta-12 desaturase gene. The location of the microsomal delta-12 desaturase gene was thus determined to be 13.6 cM proximal to locus c3838, 9.2 cM distal to locus 1At228, and 4.9 cM proximal to FadD locus on chromosome 3 [Koorneef, M. et. al. (1993) in Genetic Maps, Ed. O'Brien, S. J.; Yadav et al. (1993) Plant Physiology 103:467-476.]

EXAMPLE 5

USE OF SOYBEAN MICROSOMAL DELTA-12 DESATURASE cDNA SEQUENCE AS A RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP) MARKER

The 1.6 kb insert obtained from the plasmid pSF2-169K as previously described was radiolabelled with ^{32}P using a Random Priming Kit from Bethesda Research Laboratories under conditions recommended by the manufacturer. The resulting radioactive probe was used to probe a Southern blot (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed. (1989) Cold Spring Harbor Laboratory Press) containing genomic DNA from soybean (Glycine max (cultivar Bonus) and Glycine soja (PI81762)) digested with one of several restriction enzymes. After hybridization and washes under low stringency conditions (50 mM Tris, pH 7.5, 6X SSPE, 10% dextran sulfate, 1% SDS at 56°C for the hybridization

and initial washes, changing to 2X SSPE and 0.1% SDS for the final wash), autoradiograms were obtained, and different patterns of hybridization (polymorphisms) were identified in digests performed with restriction enzymes Hind III and Eco RI. These polymorphisms were used to map two pSF2-169k loci relative to other loci on the soybean genome essentially as described by Helentjaris et al., (Theor. Appl. Genet. (1986) 72:761-769). The map positions of the polymorphisms were determined to be in linkage group 11 between 4404.00 and 1503.00 loci (4.5 cM and 7.1 cM from 4404.00 and 1503.00, respectively) and linkage group 19 between 4010.00 and 5302.00 loci (1.9 cM and 2.7 cM from 4010.00 and 5302.00, respectively) [Rafalski, A. and Tingey, S. (1993) in Genetic Maps, Ed. O' Brien, S. J.].

EXAMPLE 6

EXPRESSION OF MICROSOMAL DELTA-12 DESATURASE IN SOYBEANS

Construction of Vectors for Transformation of *Glycine max* for Reduced Expression of Microsomal Delta-12 Desaturases in Developing Soybean Seeds

Plasmids containing the antisense *G. max* microsomal delta-12 desaturase cDNA sequence under control of the soybean Kunitz Trypsin Inhibitor 3 (KTi3) promoter (Jofuku and Goldberg, Plant Cell (1989) 1:1079-1093) , the *Phaseolus vulgaris* 7S seed storage protein (phaseolin) promoter(Sengupta-Gopalan et al., Proc. Natl. Acad. Sci. USA (1985) 82:3320-3324; Hoffman et al., Plant Mol. Biol. (1988) 11:717-729) and soybean beta-conglycinin promoter (Beachy et al., EMBO J. (1985) 4:3047-3053), were constructed. The construction of vectors expressing the soybean delta-12 desaturase antisense cDNA under the control of these promoters was facilitated by the use of the following plasmids: pML70, pCW108 and pCW109A.

The pML70 vector contains the KTi3 promoter and the KTi3 3' untranslated region and was derived from the commercially available vector pTZ18R (Pharmacia) via the intermediate plasmids pML51, pML55, pML64 and pML65. A 2.4 kb Bst BI/Eco RI fragment of the complete soybean KTi3 gene (Jofuku and Goldberg (1989) Plant Cell 1:1079-1093), which contains all 2039 nucleotides of the 5' untranslated region and 390 bases of the coding sequence of the KTi3 gene ending at the Eco RI site corresponding to bases 755 to 761 of the sequence described in Jofuku et al (1989) Plant Cell 1:427-435, was ligated into the Acc I/Eco RI sites of pTZ18R to create the plasmid pML51. The plasmid pML51 was cut with Nco I, filled in using Klenow, and religated, to destroy an Nco I site in the middle of the 5' untranslated region of the KTi3 insert, resulting in the plasmid pML55. The plasmid pML55 was partially digested with Xmn I/Eco RI to release a 0.42 kb fragment, corresponding to bases 732 to 755 of the above cited sequence, which was discarded. A synthetic Xmn I/Eco RI linker containing an Nco I site, was constructed by making a dimer of complementary synthetic oligonucleotides consisting of the coding sequence for an Xmn I site (5'-TCTTCC-3') and an Nco I site (5'-CCATGGG-3') followed directly by part of an Eco RI site (5'-GAAGG-3'). The Xmn I and Nco I/Eco RI sites were linked by a short intervening sequence (5'-ATAGCCCCCCAA-3'). This synthetic linker was ligated into the Xmn I/Eco RI sites of the 4.94 kb fragment to create the plasmid pML64. The 3' untranslated region of the KTi3 gene was amplified from the sequence described in Jofuku et al (Ibid.) by standard PCR protocols (Perkin Elmer Cetus, GeneAmp PCR kit) using the primers ML51 and ML52. Primer ML51 contained the 20 nucleotides corresponding to bases 1072 to 1091 of the above cited

sequence with the addition of nucl otides corresponding to Eco RV (5-'GATATC-3'), Nco I (5'-CCATGG-3'), Xba I (5'-TCTAGA-3'), Sma I (5'-CCCGGG-3') and Kpn I (5'-GGTACC-3') sites at the 5' end of the primer. Primer ML52 contained to the exact compliment of the nucleotides corresponding to bases 1242 to 1259 of the above cited sequence with the addition of nucleotides corresponding to Sma I (5'-CCCGGG-3'), Eco RI (5'-GAATTC-3'), Bam HI (5'-GGATCC-3') and Sal I (5'-GTCGAC-3') sites at the 5' end of the primer. The PCR-amplified 3' end of the KTi3 gene was ligated into the Nco I/Eco RI sites of pML64 to create the plasmid pML65. A synthetic multiple cloning site linker was constructed by making a dimer of complementary synthetic oligonucleotides consisting of the coding sequence for Pst I (5'-CTGCA-3'), Sal I (5'-GTCGAC-3'), Bam HI (5'-GGATCC-3') and Pst I (5'-CTGCA-3') sites. The linker was ligated into the Pst I site (directly 5' to the KTi3 promoter region) of pML65 to create the plasmid pML70.

The 1.46 kb Sma I/Kpn I fragment from pSF2-169K (soybean delta-12 desaturase cDNA described above) was ligated into the corresponding sites in pML70 resulting in the plasmid pBS10. The desaturase cDNA fragment was in the reverse (antisense) orientation with respect to the KTi3 promoter in pBS10. The plasmid pBS10 was digested with Bam HI and a 3.47 kb fragment, representing the KTi3 promoter/antisense desaturase cDNA/KTi3-3' end transcriptional unit was isolated by agarose gel electrophoresis. The vector pML18 consists of the non-tissue specific and constitutive cauliflower mosaic virus (35S) promoter (Odell et al., Nature (1985) 313:810-812; Hull et al., Virology (1987) 86:482-493), driving expression of the neomycin phosphotransferase gene described in (Beck et al. (1982) Gene 19:327-336)

followed by the 3' end of the nopaline synthase gene including nucleotides 848 to 1550 described by (Depicker et al. (1982) J. Appl. Genet. 1:561-574). This transcriptional unit was inserted into the commercial cloning vector pGEM9Z (Gibco-BRL) and is flanked at the 5' end of the 35S promoter by the restriction sites Sal I, Xba I, Bam HI and Sma I in that order. An additional Sal I site is present at the 3' end of the NOS 3' sequence and the Xba I, Bam HI and Sal I sites are unique. The 3.47 kb transcriptional unit released from pBS10 was ligated into the Bam HI site of the vector pML18. When the resulting plasmids were double digested with Sma I and Kpn I, plasmids containing inserts in the desired orientation yielded 3 fragments of 5.74, 2.69 and 1.46 kb. A plasmid with the transcriptional unit in the correct orientation was selected and was designated pBS13.

The pCW108 vector contains the bean phaseolin promoter and 3' untranslated region and was derived from the commercially available pUC18 plasmid (Gibco-BRL) via plasmids AS3 and pCW104. Plasmid AS3 contains 495 base pairs of the bean (Phaseolus vulgaris) phaseolin (7S seed storage protein) promoter starting with 5'-TGGTCTTTTGGT-3' followed by the entire 1175 base pairs of the 3' untranslated region of the same gene (see sequence descriptions in Doyle et al., (1986) J. Biol. Chem. 261:9228-9238 and Slightom et al.,(1983) Proc. Natl. Acad. Sci. USA, 80:1897-1901. Further sequence description may be found in WO 9113993) cloned into the Hind III site of pUC18. The additional cloning sites of the pUC18 multiple cloning region (Eco RI, Sph I, Pst I and Sal I) were removed by digesting with Eco RI and Sal I, filling in the ends with Klenow and religating to yield the plasmid pCW104. A new multiple cloning site was created between the 495bp of the 5'

phaseolin and the 1175bp of the 3' phaseolin by inserting a dimer of complementary synthetic oligonucleotides consisting of the coding sequence for a Nco I site (5'-CCATGG-3') followed by three filler bases (5'-TAG-3'), the coding sequence for a Sma I site (5'-CCCGGG-3'), the last three bases of a Kpn I site (5'-TAC-3'), a cytosine and the coding sequence for an Xba I site (5'-TCTAGA-3') to create the plasmid pCW108. This plasmid contains unique Nco I, Sma I, Kpn I and Xba I sites directly behind the phaseolin promoter. The 1.4 kb Eco RV/Sma I fragment from pSF2-169K was ligated into the Sma I site of the commercially available phagemid pBC SK+ (Stratagene). A phagemid with the cDNA in the desired orientation was selected by digesting with Pfl MI/Xho I to yield fragments of approx. 1 kb and 4 kb and designated pM1-SF2. The 1.4 kb Xmn I/Xba I fragment from pM1-SF2 was inserted into the Sma I/Xba I sites of pCW108 to yield the plasmid pBS11, which has the soybean delta-12 desaturase cDNA in the reverse (3'-5') orientation behind the phaseolin promoter. The plasmid pBS11 was digested with Bam HI and a 3.07 kb fragment, representing the phaseolin promoter/antisense desaturase cDNA/phaseolin 3' end transcriptional unit was isolated by agarose gel electrophoresis and ligated into the Hind III site of pML18 (described above). When the resulting plasmids were digested with Xba I, plasmids containing inserts in the desired orientation yielded 2 fragments of 8.01 and 1.18 kb. A plasmid with the transcriptional unit in the correct orientation was selected and was designated pBS14.

The vector pCW109A contains the soybean b-conglycinin promoter sequence and the phaseolin 3' untranslated region and is a modified version of vector pCW109 which was derived from the commercially available plasmid pUC18 (Gibco-BRL). The vector pCW109 was made

by inserting into the Hind III site of the cloning vector pUC18 a 555 bp 5' non-coding region (containing the promoter region) of the b-conglycinin gene followed by the multiple cloning sequence containing the restriction endonuclease sites for Nco I, Sma I, Kpn I and Xba I, as described for pCW108 above, then 1174 bp of the common bean phaseolin 3' untranslated region into the Hind III site (described above). The b-conglycinin promoter region used is an allele of the published b-conglycinin gene (Doyle et al., J. Biol. Chem. (1986) 261:9228-9238) due to differences at 27 nucleotide positions. Further sequence description of this gene may be found in Slightom (WO 9113993). To facilitate use in antisense constructions, the Nco I site and potential translation start site in the plasmid pCW109 was destroyed by digestion with Nco I, mung bean exonuclease digestion and re-ligation of the blunt site to give the modified plasmid pCW109A. The plasmid pCW109A was digested with Hind III and the resulting 1.84 kb fragment, which contained the b-conglycinin/ antisense delta-12 desaturase cDNA/phaseolin 3' untranslated region, was gel isolated. The plasmid pML18 (described above) was digested with Xba I, filled in using Klenow and religated, in order to remove the Xba I site. The resulting plasmid was designated pBS16. The 1.84 kb fragment of plasmid pCW109A (described above) was ligated into the Hind III site of pBS16. A plasmid containing the insert in the desired orientation yielded a 3.53 kb and 4.41 kb fragment when digested with Kpn I and this plasmid was designated pCST2. The Xmn I/Xba I fragment of pML1-SF2 (described above) was ligated into the Sma I/Xba I sites of pCST2 to yield the vector pST11.

Transformation Of Somatic Soybean Embryo Cultures and Regeneration Of Soybean Plants

Soybean embryogenic suspension cultures were maintained in 35 mL liquid media (SB55 or SBP6) on a rotary shaker, 150 rpm, at 28°C with mixed florescent and incandescent lights on a 16:8 h day/night schedule. Cultures were subcultured every four weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures were transformed with pCS3FdST1R by the method of particle gun bombardment (see Kline et al. (1987) Nature (London) 327:70). A DuPont Biolistic PDS1000/HE instrument (helium retrofit) was used for these transformations.

To 50 mL of a 60 mg/mL 1 mm gold particle suspension was added (in order); 5 uL DNA(1 ug/uL), 20 uL spermidine (0.1 M), and 50 ul $CaCl_2$ (2.5 M). The particle preparation was agitated for 3 min, spun in a microfuge for 10 sec and the supernatant removed. The DNA-coated particles were then washed once in 400 uL 70% ethanol and re suspended in 40 uL of anhydrous ethanol. The DNA/particle suspension was sonicated three times for 1 sec each. Five uL of the DNA-coated gold particles were then loaded on each macro carrier disk.

Approximately 300-400 mg of a four week old suspension culture was placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue were normally bombarded. Membrane rupture pressure was set at 1000 psi and the chamber was evacuated to a vacuum of 28 inches of mercury. The tissue was placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the

tissue was placed back into liquid and cultured as described above.

Eleven days post bombardment, the liquid media was exchanged with fresh SB55 containing 50 mg/mL hygromycin. The selective media was refreshed weekly. Seven weeks post bombardment, green, transformed tissue was observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue was removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Thus each new line was treated as independent transformation event. These suspensions can then be maintained as suspensions of embryos clustered in an immature developmental stage through subculture or regenerated into whole plants by maturation and germination of individual somatic embryos.

Transformed embryogenic clusters were removed from liquid culture and placed on a solid agar media (SB103) containing no hormones or antibiotics. Embryos were cultured for eight weeks at 26°C with mixed florescent and incandescent lights on a 16:8 h day/night schedule. During this period, individual embryos were removed from the clusters and analyzed at various stages of embryo development After eight weeks somatic embryos become suitable for germination. For germination, eight week old embryos were removed from the maturation medium and dried in empty petri dishes for 1 to 5 days. The dried embryos were then planted in SB71-1 medium were they were allowed to germinate under the same lighting and germination conditions described above. Germinated embryos were transferred to sterile soil and grown to maturity for seed collection.

TABLE 10

Media:

SB55 and SBP6 Stock Solutions (g/L):

MS Sulfate 100X Stock

$MgSO_4$ 7H2O	37.0
$MnSO_4$ H2O	1.69
$ZnSO_4$ 7H2O	0.86
$CuSO_4$ 5H2O	0.0025

MS Halides 100X Stock

$CaCl_2$ $2H_2O$	44.0
KI	0.083
$CoCl_2$ $6H_2O$	0.00125
KH_2PO_4	17.0
H_3BO_3	0.62
Na_2MoO_4 $2H_2O$	0.025

MS FeEDTA 100X Stock

Na_2EDTA	3.724
$FeSO_4$ $7H_2O$	2.784

B5 Vitamin Stock
- 10 g m-inositol
- 100 mg nicotinic acid
- 100 mg pyridoxine HCl
- 1 g thiamine

SB55 (per Liter)
- 10 mL each MS stocks
- 1 mL B5 Vitamin stock
- 0.8 g NH_4NO_3
- 3.033 g KNO_3
- 1 mL 2,4-D (10mg/mL stock)
- 60 g sucrose
- 0.667 g asparagine
- pH 5.7

For SBP6- substitute 0.5 mL 2,4-D

SB103 (per Liter)
- MS Salts
- 6% maltose
- 750 mg $MgCl_2$
- 0.2% Gelrite
- pH 5.7

SB71-1 (per liter)
- B5 salts
- 1ml B5 vitamin stock
- 3% sucrose
- 750mg MgCl2
- 0.2% gelrite
- pH 5.7

Analysis Of Transgenic Glycine Max Embryos and Seeds Containing An Antisense Delta-15 Desaturase: Demonstration That The Phenotype Of Transgenic Soybean Somatic Embryos Is Predictive Of The Phenotype Of Seeds Derived From Plants Regenerated From Those Embryos

While in the globular embryo state in liquid culture as described above, somatic soybean embryos contain very low amounts of triacylglycerol or storage proteins typical of maturing, zygotic soybean embryos. At this developmental stage, the ratio of total triacylglyceride to total polar lipid (phospholipids and glycolipid) is about 1:4, as is typical of zygotic soybean embryos at the developmental stage from which the somatic embryo culture was initiated. At the globular stage as well, the mRNAs for the prominent seed proteins (alpha' subunit of beta-conglycinin, Kunitz Trypsin Inhibitor 3 and Soybean Seed Lectin) are essentially absent. Upon transfer to hormone free media to allow differentiation to the maturing somatic embryo state as described above, triacylglycerol becomes the most abundant lipid class. As well, mRNAs for alpha'-subunit of beta-conglycinin, Kunitz Trypsin Inhibitor 3 and Soybean Seed Lectin become very abundant messages in the total mRNA population. In these respects the somatic soybean embryo system behaves very similarly to maturing zygotic soybean embryos in vivo, and is therefore a good and rapid model system for analyzing the phenotypic effects of modifying the expression of genes in the fatty acid biosynthesis pathway. Furthermore, the model system is predictive of the fatty acid composition of seeds from plants derived from transgenic embryos. Liquid culture globular embryos transformed with a vector containing a soybean microsomal delta-15 desaturase, in a reverse orientation and under the control of soybean conglycinin promoter

(pCS3FdST 1R), gave rise to mature embryos with a reduced 18:3 content (WO 9311245). A number of embryos from line A2872 (control tissue transformed with pCST) and from lines 299/1/3, 299/15/1, 303/7/1, 306/3/1, 306/4/3, 306/4/5 (line 2872 transformed with plasmid pCS3FdST1R) were analyzed for fatty acid content. Fatty acid analysis was performed as described in WO 9311245 using single embryos as the tissue source. Mature, somatic embryos from each of these lines were also regenerated into soybean plants by transfer to regeneration medium as described above. A number of seeds taken from plants regenerated from these embryo lines were analyzed for fatty acid content. The relative fatty-acid composition of embryos taken from tissue transformed with pCS3FdST1R was compared with relative fatty-acid composition of seeds taken from plants derived from embryos transformed with pCS3FdST1R. Also, relative fatty acid compositions of embryos and seeds transformed with pCS3FdST1R were compared with control tissue, transformed with pCST. In all cases where a reduced 18:3 content was seen in a transgenic embryo line, compared with the control, a reduced 18:3 content was also observed in segregating seeds of plants derived from that line, when compared with the control seed (Table 11).

TABLE 11

Antisense Delta-15 Desaturase:
Relative 18:3 Content Of Embryos And Seeds Of Control (A2172) And Transgenic (299-, 303-, 306-) Soybean Lines

Soybean Line	Embryo	Embryo	Seed	Seed
	av.%18:3	lowest %18:3	av.%18:3*	lowest %18:3
A2872 (control)	12.1 (2.6)	8.5	8.9 (0.8)	8.0
299/1/3	5.6 (1.2)	4.5	4.3 (1.6)	2.5
299/15/1	8.9 (2.2)	5.2	2.5 (1.8)	1.4

303/7/1	7.3 (1.1)	5.9	4.9 (1.9)	2.8
306/3/1	7.0 (1.9)	5.3	2.4 (1.7)	1.3
306/4/3	8.5 (1.9)	6.4	4.5 (2.2)	2.7
306/4/5	7.6 (1.6)	5.6	4.6 (1.6)	2.7

*Seeds which were segregating with wild-type phenotype and without a copy of the transgene are not included in these averages. The number in brackets is S.D., n=10.

Thus the Applicants conclude that an altered polyunsaturated fatty acid phenotype observed in a transgenic, mature somatic embryo line is predictive of an altered fatty acid composition of seeds of plants derived from that line.

<u>Analysis Of Transgenic Glycine Max Embryos Containing An Antisense Microsomal Delta-12 Desaturase Construct</u>

The vectors pBS13, pBS14 and pST11 contain the soybean microsomal delta-12 desaturase cDNA, in the antisense orientation, under the control of the soybean Kunitz Trypsin Inhibitor 3 (KTi3), <u>Phaseolus</u> phaseolin, and soybean beta-conglycinin promoters as described above. Liquid culture globular embryos transformed with vectors pBS13, pBS14 and pST11, gave rise to mature embryo lines as described above. Fatty acid analysis was performed as described in WO 9311245 using single, mature embryos as the tissue source. A number of embryos from line A2872 (control tissue transformed with pCST) and from line A2872 transformed with vectors pBS13, pBS14 and pST11 were analyzed for fatty acid content. About 30% of the transformed lines showed an increased 18:1 content when compared with control lines transformed with pCST described above, demonstrating that the delta-12 desaturase had been inhibited in these lines. The remaining transformed lines showed relative fatty acid compositions similar to those of the control line. The relative 18:1 content of the lines showing an increased 18:1 content was as high as 50% compared with

a maximum of 12.5% in the control embryo lines. The average 18:1 content of embryo lines which showed an increased 18:1 content was about 35% (Table 11). In all the lines showing an increased 18:1 content there was a proportional decrease in the relative 18:2 content (Table 12). The relative proportions of the other major fatty acids (16:0, 18:0 and 18:3) were similar to those of the control.

TABLE 12

Summary Of Experiment In Which Soybean Embryos Were Transformed With Plasmids Containing A Soybean Antisense Microsomal Delta-12 Desaturase cDNA

		# of lines		
	# of Vector Lines	with high 18:1	highest 18:1	av. (%) 18:1
pCST (control)	---	---	12.5	10.5
pBS13	11	4	53.5	35.9
pBS14	11	2	48.7	32.6
pST11	11	3	50.1	35.9

In Table 12 the average 18:1 of transgenics is the average of all embryos transformed with a particular vector whose relative 18:1 content is greater than two standard deviations from the highest control value (12.5). The control average is the average of ten A2872 embryos (standard deviation = 1.2). The data in Table 12 are derived from Table 13 below.

TABLE 13

Relative Fatty Acid Contents Of Embryo Lines Transformed With Plasmids Containing A Soybean Antisense Delta-12 Desaturase cDNA

Embryo Line	Relative % Fatty-Acid Content				
A2872 (control)					
#	16:0	18:0	18:1	18:2	18:3
1	11.7	3.2	11.7	52.7	16.1
2	16.4	4.0	10.8	47.1	19.3
3	17.1	3.4	8.3	48.3	20.6
4	15.3	2.7	9.4	51.1	19.0
5	15.2	3.6	10.8	51.0	17.5
6	18.6	3.9	10.9	45.8	18.1
7	14.6	3.4	12.5	52.3	16.4
8	14.2	3.5	11.2	53.9	16.7
9	15.2	3.2	9.8	49.5	16.1
10	19.0	3.8	9.6	47.4	19.0
G335/4/197 (pBS13)					
#	16:0	18:0	18:1	18:2	18:3
1	12.2	3.3	42.0	23.0	17.4
2	12.4	2.7	22.4	39.0	21.9
3	12.0	3.2	42.0	23.2	18.4
G335/4/221 (pBS13)					
#	16:0	18:0	18:1	18:2	18:3
1	12.2	2.7	30.4	36.0	17.9
2	11.5	2.4	14.3	53.4	17.6
3	13.0	2.6	15.2	47.4	19.9
4	12.0	2.6	27.4	37.9	19.1
5	11.7	2.7	25.1	42.3	15.6
6	11.7	3.4	21.6	44.3	17.8
7	12.0	2.5	11.3	53.6	20.0
8	12.0	2.5	20.8	44.1	19.5
9	11.7	2.6	25.3	39.6	18.3

G335/8/174 (pBS13)

#	16:0	18:0	18:1	18:2	18:3
1	14.1	2.1	30.3	32.1	20.3
2	14.7	2.5	5.9	40.6	34.8
3	14.3	2.4	7.3	45.2	29.8

G335/8/202 (pBS13)

#	16:0	18:0	18:1	18:2	18:3
1	11.7	1.5	30.1	32.4	23.3
2	11.4	2.3	48.5	20.6	16.1
3	12.9	2.3	46.6	17.1	19.5
4	12.7	2.6	32.0	31.1	20.5
5	12.9	1.9	41.7	23.5	18.9
6	12.3	2.6	40.1	25.6	17.9
7	11.3	2.4	53.5	16.6	14.5
8	11.4	2.5	15.5	21.7	17.8
9	10.2	2.0	45.4	23.2	18.5
10	12.8	2.2	43.2	23.5	16.9

G335/6/42 (pBS14)

#	16:0	18:0	18:1	18:2	18:3
1	13.7	2.4	38.6	28.2	15.6
2	12.6	2.3	37.6	28.8	17.2
3	11.7	3.0	48.7	21.1	14.6

G335/6/104 (pBS14)

#	16:0	18:0	18:1	18:2	18:3
1	13.8	2.5	30.5	35.4	16.0
2	12.3	2.3	14.6	53.2	16.4
3	12.7	2.6	27.1	36.6	20.0
4	12.6	2.2	32.1	34.9	17.4
5	12.7	2.6	23.2	41.2	19.3
6	12.6	2.2	11.7	52.5	20.1
7	13.3	2.1	23.3	41.2	18.4

G335/1/25 (pST11)

#	16:0	18:0	18:1	18:2	18:3
1	13.7	2.8	50.7	17.5	12.1
2	14.5	3.0	41.8	23.5	15.0

3	13.9	2.9	49.1	16.8	13.6
4	12.3	2.8	47.5	19.3	14.8
G335/2/7/1 (pST11)					
#	16:0	18:0	18:1	18:2	18:3
1	15.5	4.3	21.8	38.0	17.5
2	17.8	4.1	22.0	39.5	14.0
3	15.2	3.0	20.5	42.2	16.5
G335/2/118 (pST11)					
#	16:0	18:0	18:1	18:2	18:3
1	14.1	2.7	44.7	22.6	14.0
2	15.8	2.8	37.7	26.9	14.8
3	17.3	3.4	23.3	37.9	16.0

N.B. All other transformed embryos (24 lines) had fatty acid profiles similar to those of the control.

One of these embryo lines, G335/1/25, had an average 18:2 content of less than 20% and an average 18:1 content greater than 45% (and as high as 53.5%). The Applicants expect, based on the data in table ?, that seeds derived from plants regenerated from such lines will have an equivalent or greater increase in 18:1 content and an equivalent or greater increase decrease in 18:2 content.

EXAMPLE

EXPRESSION OF MICROSOMAL DELTA-12 DESATURASE IN CANOLA

Construction Of Vectors For Transformation of Brassica Napus For Reduced Expression of Microsomal Delta-12 Desaturases in Developing Canola Seeds

An extended poly A tail was removed from the canola delta-12 desaturase sequence contained in plasmid pCF2-165D and additional restriction sites for cloning were introduced as follows. A PCR primer was synthesized corresponding to bases 354 through 371 of SEQ ID NO:3. The second PCR primer was synthesized as

the complement to bases 1253 through 1231 with 15 additional bases (GCAGATATCGCGGCC) added to the 5' end. The additonal bases encode both an EcoRV site and a NotI site. pCF2-165D was used as the template for PCR amplification using these primers. The 914 base pair product of PCR amplification was digested with EcoRV and PflMI to give an 812 base pair product corresponding to bases 450 through 1253 of pCF2-165D with the added NotI site.

pCF2-165D was digested with PstI, the PstI overhang was blunted with Klenow fragment and then digested with PflMI. The 3.5 kB fragment corresponding to pBluescript along with the 5' 450 bases of the canola Fad2 cDNA was gel purified and ligated to the above described 812 base pair fragment. The ligation product was amplified by transformation of E. coli and plasmid DNA isolation. The EcoRI site remaining at the cloning junction between pBluescript and the canola Fad2 cDNA was destroyed by digestion, blunting and religation. The recovered plasmid was called pM2CFd2.

pM2CFd2 was digested with EcoRV and SmaI to remove the Fad2 insert as a blunt ended fragment. The fragment was gel purified and cloned into the SmaI site of pBC (Stratagene, La Jolla, CA). A plasmid with the NotI site introduced by PCR oriented away from the existing NotI site in pBC was identified by NotI digestion and gel fractionation of the digests. The resulting construct then had NotI sites at both ends of the canola Fad2 cDNA fragment and was called pM3CFd2.

Vectors for transformation of the antisense cytoplasmic delta-12 desaturase constructions under control of the ß-conglycinin, Kunitz trypsin inhibitor III, napin and phaseolin promoters into plants using Agrobacterium tumefaciens were produced by constructing a binary Ti plasmid vector system (Bevan, (1984) Nucl.

Acids Res. 12:8711-8720). One starting vector for the system, (pZS199) is based on a vector which contains: (1) the chimeric gene nopaline synthase/neomycin phosphotransferase as a selectable marker for transformed plant cells (Brevan et al. (1984) Nature 304: 184-186), (2) the left and right borders of the T-DNA of the Ti plasmid (Brevan et al. (1984) Nucl. Acids Res. 12:8711-8720), (3) the E. coli lacZ a-complementing segment (Vieria and Messing (1982) Gene 19:259-267) with unique restriction endonuclease sites for Eco RI, Kpn I, Bam HI, and Sal I, (4) the bacterial replication origin from the Pseudomonas plasmid pVS1 (Itoh et al. (1984) Plasmid 11:206-220), and (5) the bacterial neomycin phosphotransferase gene from Tn5 (Berg et al. (1975) Proc. Natnl. Acad. Sci. U.S.A. 72:3628-3632) as a selectable marker for transformed A. tumefaciens. The nopaline synthase promoter in the plant selectable marker was replaced by the 35S promoter (Odell et al. (1985) Nature, 313:810-813) by a standard restriction endonuclease digestion and ligation strategy. The 35S promoter is required for efficient Brassica napus transformation as described below. A second vector (pZS212) was constructed by reversing the order of restriction sites in the unique site cloning region of pZS199

Canola napin promoter expression cassettes were consturcted as follows: Ten oligonucleotide primers were synthesized based upon the nucleotide sequence of napin lambda clone CGN1-2 published in European Patent Application EP 255378). The oligonucleotide sequences were:

- BR42 and BR43 corresponding to bases 1132 to 1156 (BR42) and the complement of bases 2248 to 2271 (BR43) of the sequence listed in Figure 2 of EP 255378.

- BR45 and BR46 corresponding to bases 1150 to 1170 (BR46) and the complement of bases 2120 to 2155 (BR45) of the sequence listed in Figure 2 of EP 255378. In addition BR46 had bases corresponding to a Sal I site (5'-GTCGAC-3') and a few additional bases (5'-TCAGGCCT-3') at its 5' end and BR45 had bases corresponding to a Bgl II site (5'-AGATCT-3') and two (5'-CT-3') additional bases at the 5' end of the primer,
- BR47 and BR48 corresponding to bases 2705 to 2723 (BR47) and bases 2643 to 2666 (BR48) of the sequence listed in Figure 2 of EP 255378. In addition BR47 had two (5'-CT-3') additional bases at the 5' end of the primer followed by bases corresponding to a Bgl II site (5'-AGATCT-3') followed by a few additional bases (5'-TCAGGCCT-3'),
- BR49 and BR50 corresponding to the complement of bases 3877 to 3897 (BR49) and the complement of bases 3985 to 3919 (BR50) of the sequence listed in Figure 2 of EP 255378. In addition BR49 had bases corresponding to a Sal I site (5'-GTCGAC-3') and a few additional bases (5'-TCAGGCCT-3') at its 5' end,
- BR57 and BR58 corresponding to the complement of bases 3875 to 3888 (BR57) and bases 2700 to 2714 (BR58) of the sequence listed in Figure 2 of EP 255378. In addition the 5' end of BR57 had some extra bases (5'-CCATGG-3') followed by bases corresponding to a Sac I site (5'-GAGCTC-3') followed by more additional bases (5'-GTCGACGAGG-3'). The 5' end of BR58 had additional bases (5'-GAGCTC-3') followed by bases corresponding to a Nco I site (5'-CCATGG-3') followed by additional bases (5'-AGATCTGGTACC-3').
- BR61 and BR62 corresponding to bases 1846 to 1865 (BR61) and bases 2094 to 2114 (BR62) of the sequence listed in Figure 2 of EP 255378. In addition the 5'

end of BR 62 had additional bases (5'-GACA-3') followed by bases corresponding to a Bgl II site (5'-AGATCT-3') followed by a few additional bases (5'-GCGGCCGC-3').

Genomic DNA from the canola variety 'Hyola401' (Zeneca Seeds) was used as a template for PCR amplification of the napin promoter and napin terminator regions. The promoter was first amplified using primers BR42 and BR43, and reamplified using primers BR45 and BR46. Plasmid pIMC01 was derived by digestion of the 1.0 kb promoter PCR product with SalI/BglII and ligation into SalI/BamHI digested pBluescript SK^+ (Stratagene). The napin terminator region was amplified using primers BR48 and BR50, and reamplified using primers BR47 and BR49. Plasmid pIMC06 was derived by digestion of the 1.2 kb terminator PCR product with SalI/BglII and ligation into SalI/BglII digested pSP72 (Promega). Using pIMC06 as a template, the terminator region was reamplified by PCR using primer BR57 and primer BR58. Plasmid pIMC101 containing both the napin promoter and terminator was generated by digestion of the PCR product with SacI/NcoI and ligation into SacI/NcoI digested pIMC01. Plasmid pIMC101 contains a 2.2 kb napin expression cassette including complete napin 5' and 3' non-translated sequences and an introduced NcoI site at the translation start ATG. Primer BR61 and primer BR62 were used to PCR amplify an ~270 bp fragment from the 3' end of the napin promoter. Plasmid pIMC401 was obtained by digestion of the resultant PCR product with EcoRI/BglII and ligation into EcoRI/BglII digested pIMC101. Plasmid pIMC401 contains a 2.2 kb napin expression cassette lacking the napin 5' non-translated sequence and includes a NotI site at the transcription start.

To construct the antisense expression vector, pM3CFd2 was digested with NotI as was pIMC401. The delta-12 desaturase containing insert from the digest of pM3CFd2 was gel isolated and ligated into the NotI digested and phosphatase treated pIMC401. An isolate in which the delta-12 desaturase was oriented antisense to the napin promoter was selected by digestion with XhoI and PflMI to give plasmid pNCFd2R. pNCFd2R was digested with SalI, phosphatase treated and ligated into pZS212 which had been opened by the same treatment. A plasmid with desired orientation of the introduced napin:delta-12 desaturase antisense transcription unit relative to the selectable marker was chosen by digestion with PvuI and the resulting binary vector was given the name pZNCFd2R.

Plasmid pML70 (described in Example 6 above) was digested with NcoI, blunted then digested with KpnI. Plasmid pM2CFd was digested with KpnI and SmaI and the isolated fragment ligated into the opened pML70 to give the antisense expression cassette pMKCFd2R. The promotor:delta-12 desaturase:terminator sequence was removed from pMKCFd2R by BamHI digestion and ligated into pZS199 which had been BamHI digested and phosphatase treated. The desired orientation relative to the selectable marker was determined by digestion with XhoI and PflMI to give the expression vector pZKCFd2R.

The expression vector containing the ß-conglicinin promoter was constructed by SmaI and EcoRV digestion of pM2CFd2 and ligation into SmaI cut pML109A. An isolate with the antisense orientation was identified by digestion with XhoI and PflmI, and the transcription unit was isolated by SalI and EcoRI digestion. The isolated SalI-EcoRI fragment was ligated into EcoRI-SalI digested pZS199 to give pCCFd2R.

The expression vector containing the phaseolin promoter was obtained using the same proceedure with pCW108 as the starting, promoter containing vector and pZS212 as the binary portion of the vector to give pZPhCFd2R.

Agrobacterium-Mediated Transformation Of Brassica Napus

The binary vectors pZNCFd2R, pZCCFd2R, pZPhCFd2R, and pZNCFd2R were transferred by a freeze/thaw method (Holsters et al. (1978) Mol Gen Genet 163:181-187) to the *Agrobacterium* strain LBA4404/pAL4404 (Hockema et al. (1983), Nature 303:179-180).

Brassica napus cultivar "Westar" was transformed by co-cultivation of seedling pieces with disarmed *Agrobacterium tumefaciens* strain LBA4404 carrying the the appropriate binary vector.

B. napus seeds were sterilized by stirring in 10% Chlorox, 0.1% SDS for thirty min, and then rinsed thoroughly with sterile distilled water. The seeds were germinated on sterile medium containing 30 mM $CaCl_2$ and 1.5% agar, and grown for six days in the dark at 24°C.

Liquid cultures of *Agrobacterium* for plant transformation were grown overnight at 28°C in Minimal A medium containing 100 mg/L kanamycin. The bacterial cells were pelleted by centrifugation and resuspended at a concentration of 10^8 cells/mL in liquid Murashige and Skoog Minimal Organic medium containing 100 µM acetosyringone.

B. napus seedling hypocotyls were cut into 5 mm segments which were immediately placed into the bacterial suspension. After 30 min, the hypocotyl pieces were removed from the bacterial suspension and placed onto BC-35 callus medium containing 100 µM acetosyringone. The plant tissue and *Agrobacteria* were co-cultivated for three days at 24°C in dim light.

The co-cultivation was terminated by transferring the hypocotyl pieces to BC-35 callus medium containing 200 mg/L carbenicillin to kill the Agrobacteria, and 25 mg/L kanamycin to select for transformed plant cell growth. The seedling pieces were incubated on this medium for three weeks at 28°C under continuous light.

After four weeks, the segments were transferred to BS-48 regeneration medium containing 200 mg/L carbenicillin and 25 mg/L kanamycin. Plant tissue was subcultured every two weeks onto fresh selective regeneration medium, under the same culture conditions described for the callus medium. Putatively transformed calli grew rapidly on regeneration medium; as calli reached a diameter of about 2 mm, they were removed from the hypocotyl pieces and placed on the same medium lacking kanamycin.

Shoots began to appear within several weeks after transfer to BS-48 regeneration medium. As soon as the shoots formed discernable stems, they were excised from the calli, transferred to MSV-1A elongation medium, and moved to a 16:8 h photoperiod at 24°C.

Once shoots had elongated several internodes, they were cut above the agar surface and the cut ends were dipped in Rootone. Treated shoots were planted directly into wet Metro-Mix 350 soiless potting medium. The pots were covered with plastic bags which were removed when the plants were clearly growing -- after about ten days.

Plants were grown under a 16:8 h photoperiod, with a daytime temperature of 23°C and a nighttime temperature of 17°C. When the primary flowering stem began to elongate, it was covered with a mesh pollen-containment bag to prevent outcrossing. Self-pollination was facilitated by shaking the plants several times each day. Fifty-one plants have thus far been obtained from transformations using both pZCCFd2R and pZPhCFd2R, 40

plants have been obtained from pZKCFd2R and 26 from pZNCFd2R.

Minimal A Bacterial Growth Medium

Dissolve in distilled water:

10.5 grams potassium phosphate, dibasic

4.5 grams potassium phosphate, monobasic

1.0 gram ammonium sulfate

0.5 gram sodium citrate, dihydrate

Make up to 979 mL with distilled water

Autoclave

Add 20 mL filter-sterilized 10% sucrose

Add 1 mL filter-sterilized 1 M $MgSO_4$

Brassica Callus Medium BC-35

Per liter:

Murashige and Skoog Minimal Organic Medium (MS salts, 100 mg/L i-inositol, 0.4 mg/L thiamine; GIBCO #510-3118)

30 grams sucrose

18 grams mannitol

0.5 mg/L 2,4-D

0.3 mg/L kinetin

0.6% agarose

pH 5.8

Brassica Regeneration Medium BS-48

Murashige and Skoog Minimal Organic Medium

Gamborg B5 Vitamins (SIGMA #1019)

10 grams glucose

250 mg xylose

600 mg MES

0.4% agarose

pH 5.7

Filter-sterilize and add after autoclaving:

2.0 mg/L zeatin

0.1 mg/L IAA

Brassica Shoot Elongation Medium MSV-1A

Murashige and Skoog Minimal Organic Medium
Gamborg B5 Vitamins
10 grams sucrose
0.6% agarose
pH 5.8

Analysis Of Transgenic Brassica Napus Seeds Containing An Antisense Microsomal Delta-12 Desaturase Construct

Fifty-one plants were obtained from transformation with both pZPhCFd2R and pZCCFd2R, 40 were obtained from pZKCFd2R, and 26 from pZNCFd2R. The relative levels of oleate (18:1), linoleate (18:2) and linolinate (18:3) change during development so that reliable determination of seed fatty acid phenotype is best obtained from seed which has undergone nomal maturation and drydown. Relatively few transformed plants have gone through to maturity, however seeds were sampled from plants which had been transferred to pots for at least 80 days and which had pods that had yellowed and contained seeds with seed coats which had black pigmentation. Plants were chosen for early anlaysis based on promotor type, presence and copy number of the inserted delta-12 desaturase antisense gene and fertility of the plant.

Fatty acid analysis was done on either individual seeds from transformed and control plants, or on 40 mg of bulk seed from individual plants as described in Example 6. Southern analysis for detection of the presence of canola delta-12 desaturase antisense genes was done on DNA obtained from leaves of transformed plants. DNA was digested either to release the promotor:delta-12 desaturase fragment from the transformation vector or to cut outside the coding region of the delta-12 desaturase antisense gene, but within the left and right T-DNA borders of the vector.

TABLE 14

Relative Fatty Acid Profiles of Microsomal Delta-12 Desaturase Antisense Transformed and Control Brassica Napus Seeds

				% of TOTAL FATTY ACIDS				
PLANT #	PROMOTER	COPY#	AGE*	16:0	18:0	18:1	18:2	18:3
Westar	control	none	82	4.6	1.2	64.6	20.9	6.6
151-22	phaseolin	>8	82	4.4	1.0	76.6	10.0	6.2
158-8	napin	1	83	3.5	1.5	81.3	6.3	4.6
westar	control	none	106	4.1	1.7	64.4	19.9	7.1
151-22	phaseolin	>8	106	4.2	1.9	74.4	9.9	6.3
151-127	phaseolin	0	106	4.1	2.3	68.4	16.9	5.2
151-268	phaseolin	1	106	4.2	2.7	73.3	12.0	4.2
153-83	conglycinin	2	106	4.1	1.6	68.5	16.7	6.3

*Seed sampeling date in days after the plant was tranferred to soil

The expected fatty acid phenotype for antisense suppression of the delta-12 desaturase is decreased relative content of 18:2 with a corresponding increase in 18:1. Plant numbers 151-22 and 158-8 both show a substantial decrease in 18:2 content of bulk seed when compared to the westar control at 83 days after planting. Plant 151-22 also shows this difference at maturity in comparison to either the westar control or plant 151-127, which was transformed with the selectable marker gene but not the delta-12 desaturase antisense gene.

Since the fatty acid analysis was done on seeds from the primary transformant, individual seed should be segregating for the presense of the transgene copy or copies. The segregating phenotypes serve as an internal control for the effect of the delta-12 desaturase antisense gene. The relative fatty acid phenotypes for 10 individual westar seeds, 10 individual 151-22 seeds and 12 individual 158-8 seeds are given in Table 15 below.

TABLE 15

Relative Fatty Acid Profiles for Individual Seeds of Control and Genetically Segregating Delta-12 Desaturase Transformed Brassica Napus Seeds

westar control

16:0	18:0	18:1	18:2	18:3
4.65	1.05	63.45	21.31	7.29
4.65	1.37	65.41	20.72	6.18
3.86	1.31	62.19	22.50	8.18
4.46	1.41	66.81	19.40	5.63
4.76	1.30	61.90	22.39	7.65
4.59	1.10	64.77	20.62	6.56
4.61	1.16	68.66	18.20	5.07
4.71	1.26	67.28	19.32	5.18
4.67	0.98	61.96	22.93	7.61
4.73	1.33	63.85	21.65	6.23

151-22

16:0	18:0	18:1	18:2	18:3
4.56	1.08	73.40	12.40	7.60
4.25	1.20	77.90	10.00	5.40
4.40	1.00	76.90	10.10	5.90
4.40	0.94	77.40	9.40	6.10
4.50	1.00	73.60	11.30	7.90
4.60	0.98	75.40	10.50	6.50
4.49	0.96	76.70	9.90	6.00
4.20	1.10	77.20	9.70	5.50
4.20	1.00	80.00	7.90	4.90
4.50	1.00	78.00	8.80	5.80

158-8

16:0	18:0	18:1	18:2	18:3
3.62	1.67	84.45	3.60	3.73
3.46	1.64	85.56	3.02	3.36
3.48	1.61	83.64	4.43	4.21
3.53	1.40	83.80	4.41	4.36
3.48	1.39	83.66	4.35	4.44

3.80	1.50	68.17	16.57	7.56
3.41	1.40	83.76	4.38	4.40
3.49	1.29	82.77	5.16	4.60
3.77	1.39	69.47	16.40	6.54
3.44	1.36	83.86	4.49	4.27
3.48	1.38	83.15	4.91	4.53
3.55	1.92	83.69	4.20	3.70

The westar control shows comparatively little seed to seed variation in content of 18:1 or 18:2. Further the ratio of 18:3/18:2 remains very constant between seeds at about 0.35. Plant #158-8 should show a segregation ratio of either 1:2:1 or 1:3 since by Southern analysis it contains a single transgene. The 1:2:1 ratio would indicate a semi-dominant, copy number effect while the 1:3 ratio would indicate complete dominance. Two wild type 158-8 segregants are clear in Table 15, while the remaing seeds may either be the same, or the two seeds at greater than 84% 18:1 may represent the homozygous transgeneic. In either case the fatty acid phenotypes of the seeds are as expected for effective delta-12 desaturase suppression in this generation. The fatty acid phenotypes of the seeds of plant 151-22 show variation in their 18:1 and 18:2 content, with 18:1 higher than the control average and 18:2 lower. The segregation is apparently quite complex, as would be expected of a multi-copy transgenic plant.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY

(ii) TITLE OF INVENTION: GENES FOR MICROSOMAL FATTY ACID DELTA-12 DESATURASES AND RELATED ENZYMES FROM PLANTS

(iii) NUMBER OF SEQUENCES: 17

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
(B) STREET: 1007 MARKET STREET
(C) CITY: WILMINGTON
(D) STATE: DELAWARE
(E) COUNTRY: U.S.A.
(F) ZIP: 19898

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: MacIntosh
(C) OPERATING SYSTEM: MacIntosh System, 6.0
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: BB-1043-A
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: U.S. 07/977,339
(B) FILING DATE: 17-NOV-1992

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Morrissey, Bruce W
(B) REGISTRATION NUMBER: 330,663
(C) REFERENCE/DOCKET NUMBER: BB-1043-A

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (302) 992-4927
(B) TELEFAX: (302) 892-7949
(C) TELEX: 835420

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1372 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Arabidopsis thaliana

(vii) IMMEDIATE SOURCE:

(B) CLONE: p92103

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 93..1244

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
AGAGAGAGAG ATTCTGCGGA GGAGCTTCTT CTTCGTAGGG TGTTCATCGT TATTAACGTT      60

ATCGCCCCTA CGTCAGCTCC ATCTCCAGAA AC ATG GGT GCA GGT GGA AGA ATG       113
                                    Met Gly Ala Gly Gly Arg Met
                                     1               5

CCG GTT CCT ACT TCT TCC AAG AAA TCG GAA ACC GAC ACC ACA AAG CGT       161
Pro Val Pro Thr Ser Ser Lys Lys Ser Glu Thr Asp Thr Thr Lys Arg
        10                  15                  20

GTG CCG TGC GAG AAA CCG CCT TTC TCG GTG GGA GAT CTG AAG AAA GCA       209
Val Pro Cys Glu Lys Pro Pro Phe Ser Val Gly Asp Leu Lys Lys Ala
    25                  30                  35

ATC CCG CCG CAT TGT TTC AAA CGC TCA ATC CCT CGC TCT TTC TCC TAC       257
Ile Pro Pro His Cys Phe Lys Arg Ser Ile Pro Arg Ser Phe Ser Tyr
 40                  45                  50                  55
```

CTT ATC AGT GAC ATC ATT ATA GCC TCA TGC TTC TAC TAC GTC GCC ACC 305
Leu Ile Ser Asp Ile Ile Ile Ala Ser Cys Phe Tyr Tyr Val Ala Thr
60 65 70

AAT TAC TTC TCT CTC CTC CCT CAG CCT CTC TCT TAC TTG GCT TGG CCA 353
Asn Tyr Phe Ser Leu Leu Pro Gln Pro Leu Ser Tyr Leu Ala Trp Pro
75 80 85

CTC TAT TGG GCC TGT CAA GGC TGT GTC CTA ACT GGT ATC TGG GTC ATA 401
Leu Tyr Trp Ala Cys Gln Gly Cys Val Leu Thr Gly Ile Trp Val Ile
90 95 100

GCC CAC GAA TGC GGT CAC CAC GCA TTC AGC GAC TAC CAA TGG CTG GAT 449
Ala His Glu Cys Gly His His Ala Phe Ser Asp Tyr Gln Trp Leu Asp
105 110 115

GAC ACA GTT GGT CTT ATC TTC CAT TCC TTC CTC CTC GTC CCT TAC TTC 497
Asp Thr Val Gly Leu Ile Phe His Ser Phe Leu Leu Val Pro Tyr Phe
120 125 130 135

TCC TGG AAG TAT AGT CAT CGC CGT CAC CAT TCC AAC ACT GGA TCC CTC 545
Ser Trp Lys Tyr Ser His Arg Arg His His Ser Asn Thr Gly Ser Leu
140 145 150

GAA AGA GAT GAA GTA TTT GTC CCA AAG CAG AAA TCA GCA ATC AAG TGG 593
Glu Arg Asp Glu Val Phe Val Pro Lys Gln Lys Ser Ala Ile Lys Trp
155 160 165

TAC GGG AAA TAC CTC AAC AAC CCT CTT GGA CGC ATC ATG ATG TTA ACC 641
Tyr Gly Lys Tyr Leu Asn Asn Pro Leu Gly Arg Ile Met Met Leu Thr
170 175 180

GTC CAG TTT GTC CTC GGG TGG CCC TTG TAC TTA GCC TTT AAC GTC TCT 689
Val Gln Phe Val Leu Gly Trp Pro Leu Tyr Leu Ala Phe Asn Val Ser
185 190 195

GGC AGA CCG TAT GAC GGG TTC GCT TGC CAT TTC TTC CCC AAC GCT CCC 737
Gly Arg Pro Tyr Asp Gly Phe Ala Cys His Phe Phe Pro Asn Ala Pro
200 205 210 215

ATC TAC AAT GAC CGA GAA CGC CTC CAG ATA TAC CTC TCT GAT GCG GGT 785
Ile Tyr Asn Asp Arg Glu Arg Leu Gln Ile Tyr Leu Ser Asp Ala Gly
220 225 230

ATT CTA GCC GTC TGT TTT GGT CTT TAC CGT TAC GCT GCT GCA CAA GGG 833
Ile Leu Ala Val Cys Phe Gly Leu Tyr Arg Tyr Ala Ala Ala Gln Gly
235 240 245

ATG GCC TCG ATG ATC TGC CTC TAC GGA GTA CCG CTT CTG ATA GTG AAT 881
Met Ala Ser Met Ile Cys Leu Tyr Gly Val Pro Leu Leu Ile Val Asn
250 255 260

GCG TTC CTC GTC TTG ATC ACT TAC TTG CAG CAC ACT CAT CCC TCG TTG 929
Ala Phe Leu Val Leu Ile Thr Tyr Leu Gln His Thr His Pro Ser Leu
265 270 275

CCT CAC TAC GAT TCA TCA GAG TGG GAC TGG CTC AGG GGA GCT TTG GCT 977
Pro His Tyr Asp Ser Ser Glu Trp Asp Trp Leu Arg Gly Ala Leu Ala
280 285 290 295

ACC GTA GAC AGA GAC TAC GGA ATC TTG AAC AAG GTG TTC CAC AAC ATT 1025
Thr Val Asp Arg Asp Tyr Gly Ile Leu Asn Lys Val Phe His Asn Ile
300 305 310

ACA GAC ACA CAC GTG GCT CAT CAC CTG TTC TCG ACA ATG CCG CAT TAT 1073
Thr Asp Thr His Val Ala His His Leu Phe Ser Thr Met Pro His Tyr
315 320 325

AAC GCA ATG GAA GCT ACA AAG GCG ATA AAG CCA ATT CTG GGA GAC TAT 1121
Asn Ala Met Glu Ala Thr Lys Ala Ile Lys Pro Ile Leu Gly Asp Tyr
330 335 340

TAC CAG TTC GAT GGA ACA CCG TGG TAT GTA GCG ATG TAT AGG GAG GCA 1169
Tyr Gln Phe Asp Gly Thr Pro Trp Tyr Val Ala Met Tyr Arg Glu Ala
345 350 355

AAG GAG TGT ATC TAT GTA GAA CCG GAC AGG GAA GGT GAC AAG AAA GGT 1217
Lys Glu Cys Ile Tyr Val Glu Pro Asp Arg Glu Gly Asp Lys Lys Gly
360 365 370 375

GTG TAC TGG TAC AAC AAT AAG TTA TGAGCATGAT GGTGAAGAAA TTGTCGACCT 1271
Val Tyr Trp Tyr Asn Asn Lys Leu
380

TTCTCTTGTC TGTTTGTCTT TTGTTAAAGA AGCTATGCTT CGTTTTAATA ATCTTATTGT 1331

CCATTTTGTT GTGTTATGAC ATTTTGGCTG CTCATTATGT T 1372

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 383 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Gly Ala Gly Gly Arg Met Pro Val Pro Thr Ser Ser Lys Lys Ser
1 5 10 15

Glu Thr Asp Thr Thr Lys Arg Val Pro Cys Glu Lys Pro Pro Phe Ser
20 25 30

Val Gly Asp Leu Lys Lys Ala Ile Pro Pro His Cys Phe Lys Arg Ser
35 40 45

Ile Pro Arg Ser Phe Ser Tyr Leu Ile Ser Asp Ile Ile Ile Ala Ser
50 55 60

Cys Phe Tyr Tyr Val Ala Thr Asn Tyr Phe Ser Leu Leu Pro Gln Pro
65 70 75 80

Leu Ser Tyr Leu Ala Trp Pro Leu Tyr Trp Ala Cys Gln Gly Cys Val
85 90 95

Leu Thr Gly Ile Trp Val Ile Ala His Glu Cys Gly His His Ala Phe
100 105 110

Ser Asp Tyr Gln Trp Leu Asp Asp Thr Val Gly Leu Ile Phe His Ser
115 120 125

Phe Leu Leu Val Pro Tyr Phe Ser Trp Lys Tyr Ser His Arg Arg His
130 135 140

His Ser Asn Thr Gly Ser Leu Glu Arg Asp Glu Val Phe Val Pro Lys
145 150 155 160

Gln Lys Ser Ala Ile Lys Trp Tyr Gly Lys Tyr Leu Asn Asn Pro Leu
165 170 175

Gly Arg Ile Met Met Leu Thr Val Gln Phe Val Leu Gly Trp Pro Leu
180 185 190

Tyr Leu Ala Phe Asn Val Ser Gly Arg Pro Tyr Asp Gly Phe Ala Cys
195 200 205

His Phe Phe Pro Asn Ala Pro Ile Tyr Asn Asp Arg Glu Arg Leu Gln
210 215 220

Ile Tyr Leu Ser Asp Ala Gly Ile Leu Ala Val Cys Phe Gly Leu Tyr
225 230 235 240

Arg Tyr Ala Ala Ala Gln Gly Met Ala Ser Met Ile Cys Leu Tyr Gly
245 250 255

Val Pro Leu Leu Ile Val Asn Ala Phe Leu Val Leu Ile Thr Tyr Leu
260 265 270

Gln His Thr His Pro Ser Leu Pro His Tyr Asp Ser Ser Glu Trp Asp
275 280 285

Trp Leu Arg Gly Ala Leu Ala Thr Val Asp Arg Asp Tyr Gly Ile Leu
290 295 300

Asn Lys Val Phe His Asn Ile Thr Asp Thr His Val Ala His His Leu
305 310 315 320

Phe Ser Thr Met Pro His Tyr Asn Ala Met Glu Ala Thr Lys Ala Ile
325 330 335

Lys Pro Ile Leu Gly Asp Tyr Tyr Gln Phe Asp Gly Thr Pro Trp Tyr
340 345 350

Val Ala Met Tyr Arg Glu Ala Lys Glu Cys Ile Tyr Val Glu Pro Asp
355 360 365

```
Arg Glu Gly Asp Lys Lys Gly Val Tyr Trp Tyr Asn Asn Lys Leu
    370                 375                 380
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1394 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Brassica napus

(vii) IMMEDIATE SOURCE:

(B) CLONE: pCF2-165D

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 99..1250

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
GAGAGGAGAC AGAGACAGAG AGAGAGTTGA GAGAGCTCTC GTAGGTTATC GTATTAACGT      60

AATCTTCAAT CCCCCCTACG TCAGCCAGCT CAAGAAAC ATG GGT GCA GGT GGA         113
                                          Met Gly Ala Gly Gly
                                            1               5

AGA ATG CAA GTG TCT CCT CCC TCC AAA AAG TCT GAA ACC GAC AAC ATC       161
Arg Met Gln Val Ser Pro Pro Ser Lys Lys Ser Glu Thr Asp Asn Ile
                 10                  15                  20

AAG CGC GTA CCC TGC GAG ACA CCG CCC TTC ACT GTC GGA GAA CTC AAG       209
Lys Arg Val Pro Cys Glu Thr Pro Pro Phe Thr Val Gly Glu Leu Lys
             25                  30                  35

AAA GCA ATC CCA CCG CAC TGT TTC AAG CGC TCG ATC CCT CGC TCT TTC       257
Lys Ala Ile Pro Pro His Cys Phe Lys Arg Ser Ile Pro Arg Ser Phe
         40                  45                  50

TCC CAC CTC ATC TGG GAC ATC ATC ATA GCC TCC TGC TTC TAC TAC GTC       305
Ser His Leu Ile Trp Asp Ile Ile Ile Ala Ser Cys Phe Tyr Tyr Val
     55                  60                  65
```

```
GCC ACC ACT TAC TTC CCT CTC CTC CCT AAC CCT CTC TCC TAC TTC GCC     353
Ala Thr Thr Tyr Phe Pro Leu Leu Pro Asn Pro Leu Ser Tyr Phe Ala
 70                 75                  80                  85

TGG CCT CTC TAC TGG GCC TGC CAG GGC TGC GTC CTA ACC GGC GTC TGG     401
Trp Pro Leu Tyr Trp Ala Cys Gln Gly Cys Val Leu Thr Gly Val Trp
                 90                  95                 100

GTC ATA GCC CAC GAG TGC GGC CAC GCA GCC TTC AGC GAC TAC CAG TGG     449
Val Ile Ala His Glu Cys Gly His Ala Ala Phe Ser Asp Tyr Gln Trp
            105                 110                 115

CTG GAC GAC ACC GTC GGC CTC ATC TTC CAC TCC TTC CTC CTC GTC CCT     497
Leu Asp Asp Thr Val Gly Leu Ile Phe His Ser Phe Leu Leu Val Pro
        120                 125                 130

TAC TTC TCC TGG AAG TAC AGT CAT CGA CGC CAC CAT TCC AAC ACT GGC     545
Tyr Phe Ser Trp Lys Tyr Ser His Arg Arg His His Ser Asn Thr Gly
    135                 140                 145

TCC CTC GAG AGA GAC GAA GTG TTT GTC CCA AGA AGA AGT CAG ACA TCA     593
Ser Leu Glu Arg Asp Glu Val Phe Val Pro Arg Arg Ser Gln Thr Ser
150                 155                 160                 165

AGT GGT ACG GCA AGT ACC TCA ACA ACC TTT GGA CGC ACC GTG ATG TTA     641
Ser Gly Thr Ala Ser Thr Ser Thr Thr Phe Gly Arg Thr Val Met Leu
                170                 175                 180

ACG GTT CAG TTC ACT CTC GGC TGG CCT TTG TAC TTA GCC TTC AAC GTC     689
Thr Val Gln Phe Thr Leu Gly Trp Pro Leu Tyr Leu Ala Phe Asn Val
            185                 190                 195

TCG GGG AGA CCT TAC GAC GGC GGC TTC GCT TGC CAT TTC CAC CCC AAC     737
Ser Gly Arg Pro Tyr Asp Gly Gly Phe Ala Cys His Phe His Pro Asn
        200                 205                 210

GCT CCC ATC TAC AAC GAC CGT GAG CGT CTC CAG ATA TAC ATC TCC GAC     785
Ala Pro Ile Tyr Asn Asp Arg Glu Arg Leu Gln Ile Tyr Ile Ser Asp
    215                 220                 225

GCT GGC ATC CTC GCC GTC TGC TAC GGT CTG CTA CCG TAC GCT GCT GTC     833
Ala Gly Ile Leu Ala Val Cys Tyr Gly Leu Leu Pro Tyr Ala Ala Val
230                 235                 240                 245

CAA GGA GTT GCC TCG ATG GTC TGC TTC CTA CGA GTT CCT CTT CTG ATT     881
Gln Gly Val Ala Ser Met Val Cys Phe Leu Arg Val Pro Leu Leu Ile
                250                 255                 260

GTC AAC GGG TTC TTA GTT TTG ATC ACT TAC TTG CAG CAC ACG CAT CCT     929
Val Asn Gly Phe Leu Val Leu Ile Thr Tyr Leu Gln His Thr His Pro
            265                 270                 275

TCC CTG CCT CAC TAT GAC TCG TCT GAG TGG GAT TGG TTG AGG GGA GCT     977
Ser Leu Pro His Tyr Asp Ser Ser Glu Trp Asp Trp Leu Arg Gly Ala
        280                 285                 290
```

TTG GCC ACC GTT GAC AGA GAC TAC GGA ATC TTG AAC CAA GGC TTC CAC 1025
Leu Ala Thr Val Asp Arg Asp Tyr Gly Ile Leu Asn Gln Gly Phe His
295 300 305

AAT ATC ACG GAC ACG CAC GAG GCG CAT CAC CTG TTC TCG ACC ATG CCG 1073
Asn Ile Thr Asp Thr His Glu Ala His His Leu Phe Ser Thr Met Pro
310 315 320 325

CAT TAT CAT GCG ATG GAA GCT ACG AAG GCG ATA AAG CCG ATA CTG GGA 1121
His Tyr His Ala Met Glu Ala Thr Lys Ala Ile Lys Pro Ile Leu Gly
330 335 340

GAG TAT TAT CAG TTC GAT GGG ACG CCG GTG GTT AAG GCG ATG TGG AGG 1169
Glu Tyr Tyr Gln Phe Asp Gly Thr Pro Val Val Lys Ala Met Trp Arg
345 350 355

GAG GCG AAG GAG TGT ATC TAT GTG GAA CCG GAC AGG CAA GGT GAG AAG 1217
Glu Ala Lys Glu Cys Ile Tyr Val Glu Pro Asp Arg Gln Gly Glu Lys
360 365 370

AAA GGT GTG TTC TGG TAC AAC AAT AAG TTA TGAAGCAAAG AAGAAACTGA 1267
Lys Gly Val Phe Trp Tyr Asn Asn Lys Leu
375 380

ACCTTTCTCT TCTATCAATT GTCTTTGTTT AAGAAGCTAT GTTTCTGTTT CAATAATCTT 1327

AATTATCCAT TTTGTTGTGT TTTCTGACAT TTTGGCTAAA ATTATGTGAT GTTGGAAGTT 1387

AGTGTCT 1394

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 383 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Gly Ala Gly Gly Arg Met Gln Val Ser Pro Pro Ser Lys Lys Ser
1 5 10 15

Glu Thr Asp Asn Ile Lys Arg Val Pro Cys Glu Thr Pro Pro Phe Thr
20 25 30

Val Gly Glu Leu Lys Lys Ala Ile Pro Pro His Cys Phe Lys Arg Ser
35 40 45

Ile Pro Arg Ser Phe Ser His Leu Ile Trp Asp Ile Ile Ile Ala Ser
50 55 60

Cys Phe Tyr Tyr Val Ala Thr Thr Tyr Phe Pro Leu Leu Pro Asn Pro
65 70 75 80

Leu Ser Tyr Phe Ala Trp Pro Leu Tyr Trp Ala Cys Gln Gly Cys Val
85 90 95

Leu Thr Gly Val Trp Val Ile Ala His Glu Cys Gly His Ala Ala Phe
100 105 110

Ser Asp Tyr Gln Trp Leu Asp Asp Thr Val Gly Leu Ile Phe His Ser
115 120 125

Phe Leu Leu Val Pro Tyr Phe Ser Trp Lys Tyr Ser His Arg Arg His
130 135 140

His Ser Asn Thr Gly Ser Leu Glu Arg Asp Glu Val Phe Val Pro Arg
145 150 155 160

Arg Ser Gln Thr Ser Ser Gly Thr Ala Ser Thr Ser Thr Thr Phe Gly
165 170 175

Arg Thr Val Met Leu Thr Val Gln Phe Thr Leu Gly Trp Pro Leu Tyr
180 185 190

Leu Ala Phe Asn Val Ser Gly Arg Pro Tyr Asp Gly Gly Phe Ala Cys
195 200 205

His Phe His Pro Asn Ala Pro Ile Tyr Asn Asp Arg Glu Arg Leu Gln
210 215 220

Ile Tyr Ile Ser Asp Ala Gly Ile Leu Ala Val Cys Tyr Gly Leu Leu
225 230 235 240

Pro Tyr Ala Ala Val Gln Gly Val Ala Ser Met Val Cys Phe Leu Arg
245 250 255

Val Pro Leu Leu Ile Val Asn Gly Phe Leu Val Leu Ile Thr Tyr Leu
260 265 270

Gln His Thr His Pro Ser Leu Pro His Tyr Asp Ser Ser Glu Trp Asp
275 280 285

Trp Leu Arg Gly Ala Leu Ala Thr Val Asp Arg Asp Tyr Gly Ile Leu
290 295 300

Asn Gln Gly Phe His Asn Ile Thr Asp Thr His Glu Ala His His Leu
305 310 315 320

Phe Ser Thr Met Pro His Tyr His Ala Met Glu Ala Thr Lys Ala Ile
325 330 335

Lys Pro Ile Leu Gly Glu Tyr Tyr Gln Phe Asp Gly Thr Pro Val Val
340 345 350

Lys Ala Met Trp Arg Glu Ala Lys Glu Cys Ile Tyr Val Glu Pro Asp
355 360 365

```
Arg Gln Gly Glu Lys Lys Gly Val Phe Trp Tyr Asn Asn Lys Leu
    370                 375                 380
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1462 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Glycine max

(vii) IMMEDIATE SOURCE:

(B) CLONE: pSF2-165K

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 108..1247

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
CCATATACTA ATATTTGCTT GTATTGATAG CCCCTCCGTT CCCAAGAGTA TAAAACTGCA       60

TCGAATAATA CAAGCCACTA GGCATGGGTC TAGCAAAGGA AACAACA ATG GGA GGT        116
                                                    Met Gly Gly
                                                      1

AGA GGT CGT GTG GCC AAA GTG GAA GTT CAA GGG AAG AAG CCT CTC TCA        164
Arg Gly Arg Val Ala Lys Val Glu Val Gln Gly Lys Lys Pro Leu Ser
      5                  10                  15

AGG GTT CCA AAC ACA AAG CCA CCA TTC ACT GTT GGC CAA CTC AAG AAA        212
Arg Val Pro Asn Thr Lys Pro Pro Phe Thr Val Gly Gln Leu Lys Lys
 20                  25                  30                  35

GCA ATT CCA CCA CAC TGC TTT CAG CGC TCC CTC CTC ACT TCA TTC TCC        260
Ala Ile Pro Pro His Cys Phe Gln Arg Ser Leu Leu Thr Ser Phe Ser
                 40                  45                  50

TAT GTT GTT TAT GAC CTT TCA TTT GCC TTC ATT TTC TAC ATT GCC ACC        308
Tyr Val Val Tyr Asp Leu Ser Phe Ala Phe Ile Phe Tyr Ile Ala Thr
             55                  60                  65
```

```
ACC TAC TTC CAC CTC CTT CCT CAA CCC TTT TCC CTC ATT GCA TGG CCA     356
Thr Tyr Phe His Leu Leu Pro Gln Pro Phe Ser Leu Ile Ala Trp Pro
         70                  75                  80

ATC TAT TGG GTT CTC CAA GGT TGC CTT CTC ACT GGT GTG TGG GTG ATT     404
Ile Tyr Trp Val Leu Gln Gly Cys Leu Leu Thr Gly Val Trp Val Ile
     85                  90                  95

GCT CAC GAG TGT GGT CAC CAT GCC TTC AGC AAG TAC CAA TGG GTT GAT     452
Ala His Glu Cys Gly His His Ala Phe Ser Lys Tyr Gln Trp Val Asp
100                 105                 110                 115

GAT GTT GTG GGT TTG ACC CTT CAC TCA ACA CTT TTA GTC CCT TAT TTC     500
Asp Val Val Gly Leu Thr Leu His Ser Thr Leu Leu Val Pro Tyr Phe
                120                 125                 130

TCA TGG AAA ATA AGC CAT CGC CGC CAT CAC TCC AAC ACA GGT TCC CTT     548
Ser Trp Lys Ile Ser His Arg Arg His His Ser Asn Thr Gly Ser Leu
            135                 140                 145

GAC CGT GAT GAA GTG TTT GTC CCA AAA CCA AAA TCC AAA GTT GCA TGG     596
Asp Arg Asp Glu Val Phe Val Pro Lys Pro Lys Ser Lys Val Ala Trp
        150                 155                 160

TTT TCC AAG TAC TTA AAC AAC CCT CTA GGA AGG GCT GTT TCT CTT CTC     644
Phe Ser Lys Tyr Leu Asn Asn Pro Leu Gly Arg Ala Val Ser Leu Leu
    165                 170                 175

GTC ACA CTC ACA ATA GGG TGG CCT ATG TAT TTA GCC TTC AAT GTC TCT     692
Val Thr Leu Thr Ile Gly Trp Pro Met Tyr Leu Ala Phe Asn Val Ser
180                 185                 190                 195

GGT AGA CCC TAT GAT AGT TTT GCA AGC CAC TAC CAC CCT TAT GCT CCC     740
Gly Arg Pro Tyr Asp Ser Phe Ala Ser His Tyr His Pro Tyr Ala Pro
                200                 205                 210

ATA TAT TCT AAC CGT GAG AGG CTT CTG ATC TAT GTC TCT GAT GTT GCT     788
Ile Tyr Ser Asn Arg Glu Arg Leu Leu Ile Tyr Val Ser Asp Val Ala
            215                 220                 225

TTG TTT TCT GTG ACT TAC TCT CTC TAC CGT GTT GCA ACC CTG AAA GGG     836
Leu Phe Ser Val Thr Tyr Ser Leu Tyr Arg Val Ala Thr Leu Lys Gly
        230                 235                 240

TTG GTT TGG CTG CTA TGT GTT TAT GGG GTG CCT TTG CTC ATT GTG AAC     884
Leu Val Trp Leu Leu Cys Val Tyr Gly Val Pro Leu Leu Ile Val Asn
    245                 250                 255

GGT TTT CTT GTG ACT ATC ACA TAT TTG CAG CAC ACA CAC TTT GCC TTG     932
Gly Phe Leu Val Thr Ile Thr Tyr Leu Gln His Thr His Phe Ala Leu
260                 265                 270                 275

CCT CAT TAC GAT TCA TCA GAA TGG GAC TGG CTG AAG GGA GCT TTG GCA     980
Pro His Tyr Asp Ser Ser Glu Trp Asp Trp Leu Lys Gly Ala Leu Ala
                280                 285                 290
```

ACT ATG GAC AGA GAT TAT GGG ATT CTG AAC AAG GTG TTT CAT CAC ATA 1028
Thr Met Asp Arg Asp Tyr Gly Ile Leu Asn Lys Val Phe His His Ile
295 300 305

ACT GAT ACT CAT GTG GCT CAC CAT CTC TTC TCT ACA ATG CCA CAT TAC 1076
Thr Asp Thr His Val Ala His His Leu Phe Ser Thr Met Pro His Tyr
310 315 320

CAT GCA ATG GAG GCA ACC AAT GCA ATC AAG CCA ATA TTG GGT GAG TAC 1124
His Ala Met Glu Ala Thr Asn Ala Ile Lys Pro Ile Leu Gly Glu Tyr
325 330 335

TAC CAA TTT GAT GAC ACA CCA TTT TAC AAG GCA CTG TGG AGA GAA GCG 1172
Tyr Gln Phe Asp Asp Thr Pro Phe Tyr Lys Ala Leu Trp Arg Glu Ala
340 345 350 355

AGA GAG TGC CTC TAT GTG GAG CCA GAT GAA GGA ACA TCC GAG AAG GGC 1220
Arg Glu Cys Leu Tyr Val Glu Pro Asp Glu Gly Thr Ser Glu Lys Gly
360 365 370

GTG TAT TGG TAC AGG AAC AAG TAT TGATGGAGCA ACCAATGGGC CATAGTGGGA 1274
Val Tyr Trp Tyr Arg Asn Lys Tyr
375 380

GTTATGGAAG TTTTGTCATG TATTAGTACA TAATTAGTAG AATGTTATAA ATAAGTGGAT 1334

TTGCCGCGTA ATGACTTTGT GTGTATTGTG AAACAGCTTG TTGCGATCAT GGTTATAATG 1394

TAAAAATAAT TCTGGTATTA ATTACATGTG GAAAGTGTTC TGCTTATAGC TTTCTGCCTA 1454

AAAAAAAA 1462

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 379 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Gly Gly Arg Gly Arg Val Ala Lys Val Glu Val Gln Gly Lys Lys
1 5 10 15

Pro Leu Ser Arg Val Pro Asn Thr Lys Pro Pro Phe Thr Val Gly Gln
20 25 30

Leu Lys Lys Ala Ile Pro Pro His Cys Phe Gln Arg Ser Leu Leu Thr
35 40 45

Ser Phe Ser Tyr Val Val Tyr Asp Leu Ser Phe Ala Phe Ile Phe Tyr
50 55 60

Ile Ala Thr Thr Tyr Phe His Leu Leu Pro Gln Pro Phe Ser Leu Ile
65 70 75 80

Ala Trp Pro Ile Tyr Trp Val Leu Gln Gly Cys Leu Leu Thr Gly Val
85 90 95

Trp Val Ile Ala His Glu Cys Gly His His Ala Phe Ser Lys Tyr Gln
100 105 110

Trp Val Asp Asp Val Val Gly Leu Thr Leu His Ser Thr Leu Leu Val
115 120 125

Pro Tyr Phe Ser Trp Lys Ile Ser His Arg Arg His His Ser Asn Thr
130 135 140

Gly Ser Leu Asp Arg Asp Glu Val Phe Val Pro Lys Pro Lys Ser Lys
145 150 155 160

Val Ala Trp Phe Ser Lys Tyr Leu Asn Asn Pro Leu Gly Arg Ala Val
165 170 175

Ser Leu Leu Val Thr Leu Thr Ile Gly Trp Pro Met Tyr Leu Ala Phe
180 185 190

Asn Val Ser Gly Arg Pro Tyr Asp Ser Phe Ala Ser His Tyr His Pro
195 200 205

Tyr Ala Pro Ile Tyr Ser Asn Arg Glu Arg Leu Leu Ile Tyr Val Ser
210 215 220

Asp Val Ala Leu Phe Ser Val Thr Tyr Ser Leu Tyr Arg Val Ala Thr
225 230 235 240

Leu Lys Gly Leu Val Trp Leu Leu Cys Val Tyr Gly Val Pro Leu Leu
245 250 255

Ile Val Asn Gly Phe Leu Val Thr Ile Thr Tyr Leu Gln His Thr His
260 265 270

Phe Ala Leu Pro His Tyr Asp Ser Ser Glu Trp Asp Trp Leu Lys Gly
275 280 285

Ala Leu Ala Thr Met Asp Arg Asp Tyr Gly Ile Leu Asn Lys Val Phe
290 295 300

His His Ile Thr Asp Thr His Val Ala His His Leu Phe Ser Thr Met
305 310 315 320

Pro His Tyr His Ala Met Glu Ala Thr Asn Ala Ile Lys Pro Ile Leu
325 330 335

Gly Glu Tyr Tyr Gln Phe Asp Asp Thr Pro Phe Tyr Lys Ala Leu Trp
340 345 350

```
Arg Glu Ala Arg Glu Cys Leu Tyr Val Glu Pro Asp Glu Gly Thr Ser
        355                 360                 365

Glu Lys Gly Val Tyr Trp Tyr Arg Asn Lys Tyr
    370                 375
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1790 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Zea mays

(vii) IMMEDIATE SOURCE:

(B) CLONE: pFad2#1

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 165..1328

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
CGGCCTCTCC CCTCCCTCCT CCCTGCAAAT CCTGCAGACA CCACCGCTCG TTTTTCTCTC        60

CGGGACAGGA GAAAAGGGGA GAGAGAGGTG AGGCGCGGTG TCCGCCCGAT CTGCTCTGCC       120

CCGACGCAGC TGTTACGACC TCCTCAGTCT CAGTCAGGAG CAAG ATG GGT GCC GGC       176
                                                 Met Gly Ala Gly
                                                   1

GGC AGG ATG ACC GAG AAG GAG CGG GAG AAG CAG GAG CAG CTC GCC CGA       224
Gly Arg Met Thr Glu Lys Glu Arg Glu Lys Gln Glu Gln Leu Ala Arg
  5                  10                  15                  20

GCT ACC GGT GGC GCC GCG ATG CAG CGG TCG CCG GTG GAG AAG CCT CCG       272
Ala Thr Gly Gly Ala Ala Met Gln Arg Ser Pro Val Glu Lys Pro Pro
                 25                  30                  35
```

TTC ACT CTG GGT CAG ATC AAG AAG GCC ATC CCG CCA CAC TGC TTC GAG 320
Phe Thr Leu Gly Gln Ile Lys Lys Ala Ile Pro Pro His Cys Phe Glu
40 45 50

CGC TCG GTG CTC AAG TCC TTC TCG TAC GTG GTC CAC GAC CTG GTG ATC 368
Arg Ser Val Leu Lys Ser Phe Ser Tyr Val Val His Asp Leu Val Ile
55 60 65

GCC GCG GCG CTC CTC TAC TTC GCG CTG GCC ATC ATA CCG GCG CTC CCA 416
Ala Ala Ala Leu Leu Tyr Phe Ala Leu Ala Ile Ile Pro Ala Leu Pro
70 75 80

AGC CCG CTC CGC TAC GCC GCC TGG CCG CTG TAC TGG ATC GCG CAG GGG 464
Ser Pro Leu Arg Tyr Ala Ala Trp Pro Leu Tyr Trp Ile Ala Gln Gly
85 90 95 100

TGC GTG TGC ACC GGC GTG TGG GTC ATC GCG CAC GAG TGC GGC CAC CAC 512
Cys Val Cys Thr Gly Val Trp Val Ile Ala His Glu Cys Gly His His
105 110 115

GCC TTC TCG GAC TAC TCG CTC CTG GAC GAC GTG GTC GGC CTG GTG CTG 560
Ala Phe Ser Asp Tyr Ser Leu Leu Asp Asp Val Val Gly Leu Val Leu
120 125 130

CAC TCG TCG CTC ATG GTG CCC TAC TTC TCG TGG AAG TAC AGC CAC CGG 608
His Ser Ser Leu Met Val Pro Tyr Phe Ser Trp Lys Tyr Ser His Arg
135 140 145

CGC CAC CAC TCC AAC ACG GGG TCC CTG GAG CGC GAC GAG GTG TTC GTG 656
Arg His His Ser Asn Thr Gly Ser Leu Glu Arg Asp Glu Val Phe Val
150 155 160

CCC AAG AAG AAG GAG GCG CTG CCG TGG TAC ACC CCG TAC GTG TAC AAC 704
Pro Lys Lys Lys Glu Ala Leu Pro Trp Tyr Thr Pro Tyr Val Tyr Asn
165 170 175 180

AAC CCG GTC GGC CGG GTG GTG CAC ATC GTG GTG CAG CTC ACC CTC GGG 752
Asn Pro Val Gly Arg Val Val His Ile Val Val Gln Leu Thr Leu Gly
185 190 195

TGG CCG CTG TAC CTG GCG ACC AAC GCG TCG GGG CGG CCG TAC CCG CGC 800
Trp Pro Leu Tyr Leu Ala Thr Asn Ala Ser Gly Arg Pro Tyr Pro Arg
200 205 210

TTC GCC TGC CAC TTC GAC CCC TAC GGC CCC ATC TAC AAC GAC CGG GAG 848
Phe Ala Cys His Phe Asp Pro Tyr Gly Pro Ile Tyr Asn Asp Arg Glu
215 220 225

CGC GCC CAG ATC TTC GTC TCG GAC GCC GGC GTC GTG GCC GTG GCG TTC 896
Arg Ala Gln Ile Phe Val Ser Asp Ala Gly Val Val Ala Val Ala Phe
230 235 240

GGG CTG TAC AAG CTG GCG GCG GCG TTC GGG GTC TGG TGG GTG GTG CGC 944
Gly Leu Tyr Lys Leu Ala Ala Ala Phe Gly Val Trp Trp Val Val Arg
245 250 255 260

GTG TAC GCC GTG CCG CTG CTG ATC GTG AAC GCG TGG CTG GTG CTC ATC 992
Val Tyr Ala Val Pro Leu Leu Ile Val Asn Ala Trp Leu Val Leu Ile
265 270 275

ACC TAC CTG CAG CAC ACC CAC CCG TCG CTC CCC CAC TAC GAC TCG AGC 1040
Thr Tyr Leu Gln His Thr His Pro Ser Leu Pro His Tyr Asp Ser Ser
280 285 290

GAG TGG GAC TGG CTG CGC GGC GCG CTG GCC ACC ATG GAC CGC GAC TAC 1088
Glu Trp Asp Trp Leu Arg Gly Ala Leu Ala Thr Met Asp Arg Asp Tyr
295 300 305

GGC ATC CTC AAC CGC GTG TTC CAC AAC ATC ACG GAC ACG CAC GTC GCG 1136
Gly Ile Leu Asn Arg Val Phe His Asn Ile Thr Asp Thr His Val Ala
310 315 320

CAC CAC CTC TTC TCC ACC ATG CCG CAC TAC CAC GCC ATG GAG GCC ACC 1184
His His Leu Phe Ser Thr Met Pro His Tyr His Ala Met Glu Ala Thr
325 330 335 340

AAG GCG ATC AGG CCC ATC CTC GGC GAC TAC TAC CAC TTC GAC CCG ACC 1232
Lys Ala Ile Arg Pro Ile Leu Gly Asp Tyr Tyr His Phe Asp Pro Thr
345 350 355

CCT GTC GCC AAG GCG ACC TGG CGC GAG GCC GGG GAA TGC ATC TAC GTC 1280
Pro Val Ala Lys Ala Thr Trp Arg Glu Ala Gly Glu Cys Ile Tyr Val
360 365 370

GAG CCC GAG GAC CGC AAG GGC GTC TTC TGG TAC AAC AAG AAG TTC TAGCCGCCGC 1335
Glu Pro Glu Asp Arg Lys Gly Val Phe Trp Tyr Asn Lys Lys Phe
375 380 385

CGCTCGCAGA GCTGAGGACG CTACCGTAGG AATGGGAGCA GAAACCAGGA GGAGGAGACG 1395

GTACTCGCCC CAAAGTCTCC GTCAACCTAT CTAATCGTTA GTCGTCAGTC TTTTAGACGG 1455

GAAGAGAGAT CATTTGGGCA CAGAGACGAA GGCTTACTGC AGTGCCATCG CTAGAGCTGC 1515

CATCAAGTAC AAGTAGGCAA ATTCGTCAAC TTAGTGTGTC CCATGTTGTT TTTCTTAGTC 1575

GTCCGCTGCT GTAGGCTTTC CGGCGGCGGT CGTTTGTGTG GTTGGCATCC GTGGCCATGC 1635

CTGTGCGTGC GTGGCCGCGC TTGTCGTGTG CGTCTGTCGT CGCGTTGGCG TCGTCTCTTC 1695

GTGCTCCCCG TGTGTTGTTG TAAAACAAGA AGATGTTTTC TGGTGTCTTT GGCGGAATAA 1755

CAGATCGTCC GAACGAAAAA AAAAAAAAAA AAAAA 1790

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 387 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met Gly Ala Gly Gly Arg Met Thr Glu Lys Glu Arg Glu Lys Gln Glu
 1               5                  10                  15

Gln Leu Ala Arg Ala Thr Gly Gly Ala Ala Met Gln Arg Ser Pro Val
            20                  25                  30

Glu Lys Pro Pro Phe Thr Leu Gly Gln Ile Lys Lys Ala Ile Pro Pro
        35                  40                  45

His Cys Phe Glu Arg Ser Val Leu Lys Ser Phe Ser Tyr Val Val His
    50                  55                  60

Asp Leu Val Ile Ala Ala Ala Leu Leu Tyr Phe Ala Leu Ala Ile Ile
65                  70                  75                  80

Pro Ala Leu Pro Ser Pro Leu Arg Tyr Ala Ala Trp Pro Leu Tyr Trp
                85                  90                  95

Ile Ala Gln Gly Cys Val Cys Thr Gly Val Trp Val Ile Ala His Glu
            100                 105                 110

Cys Gly His His Ala Phe Ser Asp Tyr Ser Leu Leu Asp Asp Val Val
        115                 120                 125

Gly Leu Val Leu His Ser Ser Leu Met Val Pro Tyr Phe Ser Trp Lys
    130                 135                 140

Tyr Ser His Arg Arg His His Ser Asn Thr Gly Ser Leu Glu Arg Asp
145                 150                 155                 160

Glu Val Phe Val Pro Lys Lys Lys Glu Ala Leu Pro Trp Tyr Thr Pro
                165                 170                 175

Tyr Val Tyr Asn Asn Pro Val Gly Arg Val Val His Ile Val Val Gln
            180                 185                 190

Leu Thr Leu Gly Trp Pro Leu Tyr Leu Ala Thr Asn Ala Ser Gly Arg
        195                 200                 205

Pro Tyr Pro Arg Phe Ala Cys His Phe Asp Pro Tyr Gly Pro Ile Tyr
    210                 215                 220

Asn Asp Arg Glu Arg Ala Gln Ile Phe Val Ser Asp Ala Gly Val Val
225                 230                 235                 240

Ala Val Ala Phe Gly Leu Tyr Lys Leu Ala Ala Ala Phe Gly Val Trp
                245                 250                 255

Trp Val Val Arg Val Tyr Ala Val Pro Leu Leu Ile Val Asn Ala Trp
            260                 265                 270

Leu Val Leu Ile Thr Tyr Leu Gln His Thr His Pro Ser Leu Pro His
        275                 280                 285
```

Tyr Asp Ser Ser Glu Trp Asp Trp Leu Arg Gly Ala Leu Ala Thr Met
    290   295   300

Asp Arg Asp Tyr Gly Ile Leu Asn Arg Val Phe His Asn Ile Thr Asp
305   310   315   320

Thr His Val Ala His His Leu Phe Ser Thr Met Pro His Tyr His Ala
                325   330   335

Met Glu Ala Thr Lys Ala Ile Arg Pro Ile Leu Gly Asp Tyr Tyr His
            340   345   350

Phe Asp Pro Thr Pro Val Ala Lys Ala Thr Trp Arg Glu Ala Gly Glu
        355   360   365

Cys Ile Tyr Val Glu Pro Glu Asp Arg Lys Gly Val Phe Trp Tyr Asn
    370   375   380

Lys Lys Phe
385

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 673 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Ricinus communis

(vii) IMMEDIATE SOURCE:

(B) CLONE: pRF2-1C

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 1..673

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TGG GTG ATG GCG CAT GAT TGT GGG CAC CAT GCC TTC AGT GAC TAT CAA 48
Trp Val Met Ala His Asp Cys Gly His His Ala Phe Ser Asp Tyr Gln
1 5 10 15

TTG CTT GAT GAT GTA GTT GGT CTT ATC CTA CAC TCC TGT CTC CTT GTC 96
Leu Leu Asp Asp Val Val Gly Leu Ile Leu His Ser Cys Leu Leu Val
20 25 30

CCT TAT TTT TCA TGG AAA CAC AGC CAT CGC CGA CAT CAT TCC AAC ACA 144
Pro Tyr Phe Ser Trp Lys His Ser His Arg Arg His His Ser Asn Thr
35 40 45

GGG TCC CTG GAA CGG GAT GAA GTG TTT GTT CCC AAG AAG AAA TCT AGT 192
Gly Ser Leu Glu Arg Asp Glu Val Phe Val Pro Lys Lys Lys Ser Ser
50 55 60

ATC CGT TGG TAT TCC AAA TAC CTC AAC AAC CCT CCA GGT CGT ATC ATG 240
Ile Arg Trp Tyr Ser Lys Tyr Leu Asn Asn Pro Pro Gly Arg Ile Met
65 70 75 80

ACA ATT GCC GTC ACA CTT TCA CTT GGC TGG CCT CTG TAC CTA GCA TTC 288
Thr Ile Ala Val Thr Leu Ser Leu Gly Trp Pro Leu Tyr Leu Ala Phe
85 90 95

AAT GTT TCA GGC AGG CCA TAT GAT CGG TTC GCC TGC CAC TAT GAC CCA 336
Asn Val Ser Gly Arg Pro Tyr Asp Arg Phe Ala Cys His Tyr Asp Pro
100 105 110

TAT GGC CCG ATC TAC AAT GAT CGC GAG CGA ATC GAG ATA TTC ATA TCA 384
Tyr Gly Pro Ile Tyr Asn Asp Arg Glu Arg Ile Glu Ile Phe Ile Ser
115 120 125

GAT GCT GGT GTT CTT GCT GTC ACT TTT GGT CTC TAC CAA CTT GCT ATA 432
Asp Ala Gly Val Leu Ala Val Thr Phe Gly Leu Tyr Gln Leu Ala Ile
130 135 140

GCG AAG GGG CTT GCT TGG GTT GTC TGT GTA TAT GGA GTG CCA TTG TTG 480
Ala Lys Gly Leu Ala Trp Val Val Cys Val Tyr Gly Val Pro Leu Leu
145 150 155 160

GTG GTG AAT TCA TTC CTT GTT CTG ATC ACA TTT CTG CAG CAT ACT CAC 528
Val Val Asn Ser Phe Leu Val Leu Ile Thr Phe Leu Gln His Thr His
165 170 175

CCT GCA TTG CCA CAT TAT GAT TCG TCG GAG TGG GAC TGG CTA AGA GGA 576
Pro Ala Leu Pro His Tyr Asp Ser Ser Glu Trp Asp Trp Leu Arg Gly
180 185 190

GCT CTA GCA ACT GTT GAC AGA GAT TAC GGG ATC TTG AAC AAG GTG TTC 624
Ala Leu Ala Thr Val Asp Arg Asp Tyr Gly Ile Leu Asn Lys Val Phe
195 200 205

CAT AAC ATA ACG GAC ACT CAA GTA GCT CAC CAC CTT TTC ACC ATG CCC C 673
His Asn Ile Thr Asp Thr Gln Val Ala His His Leu Phe Thr Met Pro
210 215 220

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 224 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Trp Val Met Ala His Asp Cys Gly His His Ala Phe Ser Asp Tyr Gln
 1               5                  10                  15

Leu Leu Asp Asp Val Val Gly Leu Ile Leu His Ser Cys Leu Leu Val
            20                  25                  30

Pro Tyr Phe Ser Trp Lys His Ser His Arg Arg His His Ser Asn Thr
        35                  40                  45

Gly Ser Leu Glu Arg Asp Glu Val Phe Val Pro Lys Lys Lys Ser Ser
    50                  55                  60

Ile Arg Trp Tyr Ser Lys Tyr Leu Asn Asn Pro Pro Gly Arg Ile Met
65                  70                  75                  80

Thr Ile Ala Val Thr Leu Ser Leu Gly Trp Pro Leu Tyr Leu Ala Phe
                85                  90                  95

Asn Val Ser Gly Arg Pro Tyr Asp Arg Phe Ala Cys His Tyr Asp Pro
            100                 105                 110

Tyr Gly Pro Ile Tyr Asn Asp Arg Glu Arg Ile Glu Ile Phe Ile Ser
        115                 120                 125

Asp Ala Gly Val Leu Ala Val Thr Phe Gly Leu Tyr Gln Leu Ala Ile
    130                 135                 140

Ala Lys Gly Leu Ala Trp Val Val Cys Val Tyr Gly Val Pro Leu Leu
145                 150                 155                 160

Val Val Asn Ser Phe Leu Val Leu Ile Thr Phe Leu Gln His Thr His
                165                 170                 175

Pro Ala Leu Pro His Tyr Asp Ser Ser Glu Trp Asp Trp Leu Arg Gly
            180                 185                 190

Ala Leu Ala Thr Val Asp Arg Asp Tyr Gly Ile Leu Asn Lys Val Phe
        195                 200                 205

His Asn Ile Thr Asp Thr Gln Val Ala His His Leu Phe Thr Met Pro
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1369 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Ricinus communis

(vii) IMMEDIATE SOURCE:

(B) CLONE: pRF197c-42

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 184..1347

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
CGGCCGGGAT TCCGGTTTTC ACACTAATTT GCAAAAAATG CATGATTTCA CCTCAAATCA      60

AACACCACAC CTTATAACTT AGTCTTAAGA GAGAGAGAGA GAGGAGACAT TTCTCTTCTC     120

TGAGATGAGC ACTTCTCTTC CAGACATCGA AGCCTCAGGA AAGTGCTTGA GAAGAGCTTG     180

AGA ATG GGA GGT GGT GGT CGC ATG TCT ACT GTC ATA ATC AGC AAC AAC       228
    Met Gly Gly Gly Gly Arg Met Ser Thr Val Ile Ile Ser Asn Asn
      1               5                  10                  15

AGT GAG AAG AAA GGA GGA AGC AGC CAC CTG GAG CGA GCG CCG CAC ACG       276
Ser Glu Lys Lys Gly Gly Ser Ser His Leu Glu Arg Ala Pro His Thr
                 20                  25                  30

AAG CCT CCT TAC ACA CTT GGT AAC CTC AAG AGA GCC ATC CCA CCC CAT       324
Lys Pro Pro Tyr Thr Leu Gly Asn Leu Lys Arg Ala Ile Pro Pro His
             35                  40                  45

TGC TTT GAA CGC TCT TTT GTG CGC TCA TTC TCC AAT TTT GCC TAT AAT       372
Cys Phe Glu Arg Ser Phe Val Arg Ser Phe Ser Asn Phe Ala Tyr Asn
         50                  55                  60
```

```
TTC TGC TTA AGT TTT CTT TCC TAC TCG ATC GCC ACC AAC TTC TTC CCT        420
Phe Cys Leu Ser Phe Leu Ser Tyr Ser Ile Ala Thr Asn Phe Phe Pro
     65                  70                  75

TAC ATC TCT TCT CCG CTC TCG TAT GTC GCT TGG CTG GTT TAC TGG CTC        468
Tyr Ile Ser Ser Pro Leu Ser Tyr Val Ala Trp Leu Val Tyr Trp Leu
 80                  85                  90                  95

TTC CAA GGC TGC ATT CTC ACT GGT CTT TGG GTC ATC GGC CAT GAA TGT        516
Phe Gln Gly Cys Ile Leu Thr Gly Leu Trp Val Ile Gly His Glu Cys
                100                 105                 110

GGC CAT CAT GCT TTT AGT GAG TAT CAG CTG GCT GAT GAC ATT GTT GGC        564
Gly His His Ala Phe Ser Glu Tyr Gln Leu Ala Asp Asp Ile Val Gly
            115                 120                 125

CTA ATT GTC CAT TCT GCA CTT CTG GTT CCA TAT TTT TCA TGG AAA TAT        612
Leu Ile Val His Ser Ala Leu Leu Val Pro Tyr Phe Ser Trp Lys Tyr
        130                 135                 140

AGC CAT CGC CGC CAC CAT TCT AAC ATA GGA TCT CTC GAG CGA GAC GAA        660
Ser His Arg Arg His His Ser Asn Ile Gly Ser Leu Glu Arg Asp Glu
    145                 150                 155

GTG TTC GTC CCG AAA TCA AAG TCG AAA ATT TCA TGG TAT TCT AAG TAC        708
Val Phe Val Pro Lys Ser Lys Ser Lys Ile Ser Trp Tyr Ser Lys Tyr
160                 165                 170                 175

TTA AAC AAC CCG CCA GGT CGA GTT TTG ACA CTT GCT GCC ACG CTC CTC        756
Leu Asn Asn Pro Pro Gly Arg Val Leu Thr Leu Ala Ala Thr Leu Leu
                180                 185                 190

CTT GGC TGG CCT TTA TAT TTA GCT TTC AAT GTC TCT GGT AGA CCT TAC        804
Leu Gly Trp Pro Leu Tyr Leu Ala Phe Asn Val Ser Gly Arg Pro Tyr
            195                 200                 205

GAT CGC TTT GCT TGC CAT TAT GAT CCC TAT GGC CCA ATA TTT TCC GAA        852
Asp Arg Phe Ala Cys His Tyr Asp Pro Tyr Gly Pro Ile Phe Ser Glu
        210                 215                 220

AGA GAA AGG CTT CAG ATT TAC ATT GCT GAC CTC GGA ATC TTT GCC ACA        900
Arg Glu Arg Leu Gln Ile Tyr Ile Ala Asp Leu Gly Ile Phe Ala Thr
    225                 230                 235

ACG TTT GTG CTT TAT CAG GCT ACA ATG GCA AAA GGG TTG GCT TGG GTA        948
Thr Phe Val Leu Tyr Gln Ala Thr Met Ala Lys Gly Leu Ala Trp Val
240                 245                 250                 255

ATG CGT ATC TAT GGG GTG CCA TTG CTT ATT GTT AAC TGT TTC CTT GTT        996
Met Arg Ile Tyr Gly Val Pro Leu Leu Ile Val Asn Cys Phe Leu Val
                260                 265                 270

ATG ATC ACA TAC TTG CAG CAC ACT CAC CCA GCT ATT CCA CGC TAT GGC       1044
Met Ile Thr Tyr Leu Gln His Thr His Pro Ala Ile Pro Arg Tyr Gly
            275                 280                 285
```

```
TCA TCG GAA TGG GAT TGG CTC CGG GGA GCA ATG GTG ACT GTC GAT AGA      1092
Ser Ser Glu Trp Asp Trp Leu Arg Gly Ala Met Val Thr Val Asp Arg
        290                 295                 300

GAT TAT GGG GTG TTG AAT AAA GTA TTC CAT AAC ATT GCA GAC ACT CAT      1140
Asp Tyr Gly Val Leu Asn Lys Val Phe His Asn Ile Ala Asp Thr His
    305                 310                 315

GTA GCT CAT CAT CTC TTT GCT ACA GTG CCA CAT TAC CAT GCA ATG GAG      1188
Val Ala His His Leu Phe Ala Thr Val Pro His Tyr His Ala Met Glu
320                 325                 330                 335

GCC ACT AAA GCA ATC AAG CCT ATA ATG GGT GAG TAT TAC CGG TAT GAT      1236
Ala Thr Lys Ala Ile Lys Pro Ile Met Gly Glu Tyr Tyr Arg Tyr Asp
                340                 345                 350

GGT ACC CCA TTT TAC AAG GCA TTG TGG AGG GAG GCA AAG GAG TGC TTG      1284
Gly Thr Pro Phe Tyr Lys Ala Leu Trp Arg Glu Ala Lys Glu Cys Leu
            355                 360                 365

TTC GTC GAG CCA GAT GAA GGA GCT CCT ACA CAA GGC GTT TTC TGG TAC      1332
Phe Val Glu Pro Asp Glu Gly Ala Pro Thr Gln Gly Val Phe Trp Tyr
        370                 375                 380

CGG AAC AAG TAT TAAAAAAGTG TCATGTAGCC TGCCG                          1369
Arg Asn Lys Tyr
    385
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 387 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Met Gly Gly Gly Gly Arg Met Ser Thr Val Ile Ile Ser Asn Asn Ser
1               5                   10                  15

Glu Lys Lys Gly Gly Ser Ser His Leu Glu Arg Ala Pro His Thr Lys
            20                  25                  30

Pro Pro Tyr Thr Leu Gly Asn Leu Lys Arg Ala Ile Pro Pro His Cys
        35                  40                  45

Phe Glu Arg Ser Phe Val Arg Ser Phe Ser Asn Phe Ala Tyr Asn Phe
    50                  55                  60

Cys Leu Ser Phe Leu Ser Tyr Ser Ile Ala Thr Asn Phe Phe Pro Tyr
65                  70                  75                  80
```

```
Ile Ser Ser Pro Leu Ser Tyr Val Ala Trp Leu Val Tyr Trp Leu Phe
                85                  90                  95

Gln Gly Cys Ile Leu Thr Gly Leu Trp Val Ile Gly His Glu Cys Gly
            100                 105                 110

His His Ala Phe Ser Glu Tyr Gln Leu Ala Asp Asp Ile Val Gly Leu
        115                 120                 125

Ile Val His Ser Ala Leu Leu Val Pro Tyr Phe Ser Trp Lys Tyr Ser
    130                 135                 140

His Arg Arg His His Ser Asn Ile Gly Ser Leu Glu Arg Asp Glu Val
145                 150                 155                 160

Phe Val Pro Lys Ser Lys Ser Lys Ile Ser Trp Tyr Ser Lys Tyr Leu
                165                 170                 175

Asn Asn Pro Pro Gly Arg Val Leu Thr Leu Ala Ala Thr Leu Leu Leu
            180                 185                 190

Gly Trp Pro Leu Tyr Leu Ala Phe Asn Val Ser Gly Arg Pro Tyr Asp
        195                 200                 205

Arg Phe Ala Cys His Tyr Asp Pro Tyr Gly Pro Ile Phe Ser Glu Arg
    210                 215                 220

Glu Arg Leu Gln Ile Tyr Ile Ala Asp Leu Gly Ile Phe Ala Thr Thr
225                 230                 235                 240

Phe Val Leu Tyr Gln Ala Thr Met Ala Lys Gly Leu Ala Trp Val Met
                245                 250                 255

Arg Ile Tyr Gly Val Pro Leu Leu Ile Val Asn Cys Phe Leu Val Met
            260                 265                 270

Ile Thr Tyr Leu Gln His Thr His Pro Ala Ile Pro Arg Tyr Gly Ser
        275                 280                 285

Ser Glu Trp Asp Trp Leu Arg Gly Ala Met Val Thr Val Asp Arg Asp
    290                 295                 300

Tyr Gly Val Leu Asn Lys Val Phe His Asn Ile Ala Asp Thr His Val
305                 310                 315                 320

Ala His His Leu Phe Ala Thr Val Pro His Tyr His Ala Met Glu Ala
                325                 330                 335

Thr Lys Ala Ile Lys Pro Ile Met Gly Glu Tyr Tyr Arg Tyr Asp Gly
            340                 345                 350

Thr Pro Phe Tyr Lys Ala Leu Trp Arg Glu Ala Lys Glu Cys Leu Phe
        355                 360                 365

Val Glu Pro Asp Glu Gly Ala Pro Thr Gln Gly Val Phe Trp Tyr Arg
    370                 375                 380
```

Asn Lys Tyr
385

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION: 1..23
(D) OTHER INFORMATION: /product= "synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TGGGTATGCC AYGANTGYGG NCA 23

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) TRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION: 1..22

(D) OTHER INFORMATION: /product= "synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AAARTGRTGG CACRTGNGTR TC 22

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2973 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Arabidopsis thaliana

(vii) IMMEDIATE SOURCE:

(B) CLONE: pAGF2-6

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 433..520

(ix) FEATURE:

(A) NAME/KEY: intron
(B) LOCATION: 521..1654

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

ATTCGGTAAT TCCTACATAT TTTAGAGATT AGTTTGAGTT TCCATCCATA CTTTACTAGT 60

GATTATAAAT TTAAAATACG TACTTTTCGA CTATAAAGTG AAACTAAGTA AATTAGAACG 120

TGATATTAAA AAGTTAATGT TCACTGTTAT ATTTTTTTCA CAAGTAAAAA ATGGGTTATT 180

TGCGGTAAAT AAAAATACCA GATATTTTGA ATTGATTAAA AAGGTTGAAA TAAGAGAGGA 240

GGGGAAAGAA AAGAAGGTGG GGGCCCAGTA TGAAAGGGAA AGGTGTCATC AAATCATCTC 300

```
TCTCTCTCTC TACCTTCGAC CCACGGGCCG TGTCCATTTA AAGCCCTGTC TCTTGCCATT   360
CCCCATCTGA CCACCAGAAG AAGAGCCACA CACTCACAAA TTAAAAAGAG AGAGAGAGAG   420
AGAGAGACAG AGAGAGAGAG AGATTCTGCG GAGGAGCTTC TTCTTCGTAG GGTGTTCATC   480
GTTATTAACG TTATCGCCCC TACGTCAGCT CCATCTCCAG GTCCGTCGCT TCTCTTCCAT   540
TTCTTCTCAT TTTCGATTTT GATTCTTATT TCTTTCCAGT AGCTCCTGCT CTGTGAATTT   600
CTCCGCTCAC GATAGATCTG CTTATACTCC TTACATTCAA CCTTAGATCT GGTCTCGATT   660
CTCTGTTTCT CTGTTTTTTT CTTTTGGTCG AGAATCTGAT GTTTGTTTAT GTTCTGTCAC   720
CATTAATAAT GATGAACTCT CTCATTCATA CAATGATTAG TTTCTCTCGT CTACCAAACG   780
ATATGTTGCA TTTTCACTTT TCTTCTTTTT TTCTAAGATG ATTTGCTTTG ACCAATTTGT   840
TTAGATCTTT ATTTTATTTT ATTTTCTGGT GGGTTGGTGG AAATTGAAAA AAAAAAAAAA   900
AAAAGCATAA ATTGTTATTT GTTAATGTAT TCATTTTTTG GCTATTTGTT CTGGGTAAAA   960
ATCTGCTTCT ACTGTTGAAT CTTTCCTGGA TTTTTTACTC CTATTGGGTT TTTATAGTAA  1020
AAATACATAA TAAAAGGAAA ACAAAAGTTT TATAGATTCT CTTAAACCCC TTACGATAAA  1080
AGTTGGAATC AAAATAATTC AGGATCAGAT GCTCTTTGAT TGATTCAGAT GCGATTACAG  1140
TTGCATGGAA AATTTTCTAG ATCCGTCGTC ACATTTTATT TTCTGTTTAA ATATCTAAAT  1200
CTGATATATG ATGTCGACAA ATTCTGGTGG CTTATACATC ACTTCAACTG TTTTCTTTTG  1260
GCTTTGTTTG TCAACTTGGT TTTCAATACG ATTTGTGATT TCGATCGCTG AATTTTTAAT  1320
ACAAGCAAAC TGATGTTAAC CACAAGCAAG AGATGTGACC TGCCTTATTA ACATCGTATT  1380
ACTTACTACT AGTCGTATTC TCAACGCAAT CGTTTTTGTA TTTCTCACAT TATGCCGCTT  1440
CTCTACTCTT TATTCCTTTT GGTCCACGCA TTTTCTATTT GTGGCAATCC CTTTCACAAC  1500
CTGATTTCCC ACTTTGGATC ATTTGTCTGA AGACTCTCTT GAATCGTTAC CACTTGTTTC  1560
TTGTGCATGC TCTGTTTTTT AGAATTAATG ATAAAACTAT TCCATAGTCT TGAGTTTTCA  1620
GCTTGTTGAT TCTTTTGCTT TTGGTTTTCT GCAGAAACAT GGGTGCAGGT GGAAGAATGC  1680
CGGTTCCTAC TTCTTCCAAG AAATCGGAAA CCGACACCAC AAAGCGTGTG CCGTGCGAGA  1740
AACCGCCTTT CTCGGTGGGA GATCTGAAGA AAGCAATCCC GCCGCATTGT TTCAAACGCT  1800
CAATCCCTCG CTCTTTCTCC TACCTTATCA GTGACATCAT TATAGCCTCA TGCTTCTACT  1860
ACGTCGCCAC CAATTACTTC TCTCTCCTCC CTCAGCCTCT CTCTTACTTG GCTTGGCCAC  1920
TCTATTGGGC CTGTCAAGGC TGTGTCCTAA CTGGTATCTG GGTCATAGCC CACGAATGCG  1980
GTCACCACGC ATTCAGCGAC TACCAATGGC TGGATGACAC AGTTGGTCTT ATCTTCCATT  2040
```

CCTTCCTCCT CGTCCCTTAC TTCTCCTGGA AGTATAGTCA TCGCCGTCAC CATTCCAACA 2100

CTGGATCCCT CGAAAGAGAT GAAGTATTTG TCCCAAAGCA GAAATCAGCA ATCAAGTGGT 2160

ACGGGAAATA CCTCAACAAC CCTCTTGGAC GCATCATGAT GTTAACCGTC CAGTTTGTCC 2220

TCGGGTGGCC CTTGTACTTA GCCTTTAACG TCTCTGGCAG ACCGTATGAC GGGTTCGCTT 2280

GCCATTTCTT CCCCAACGCT CCCATCTACA ATGACCGAGA ACGCCTCCAG ATATACCTCT 2340

CTGATGCGGG TATTCTAGCC GTCTGTTTTG GTCTTTACCG TTACGCTGCT GCACAAGGGA 2400

TGGCCTCGAT GATCTGCCTC TACGGAGTAC CGCTTCTGAT AGTGAATGCG TTCCTCGTCT 2460

TGATCACTTA CTTGCAGCAC ACTCATCCCT CGTTGCCTCA CTACGATTCA TCAGAGTGGG 2520

ACTGGCTCAG GGGAGCTTTG GCTACCGTAG ACAGAGACTA CGGAATCTTG AACAAGGTGT 2580

TCCACAACAT TACAGACACA CACGTGGCTC ATCACCTGTT CTCGACAATG CCGCATTATA 2640

ACGCAATGGA AGCTACAAAG GCGATAAAGC CAATTCTGGG AGACTATTAC CAGTTCGATG 2700

GAACACCGTG GTATGTGGCG ATGTATAGGG AGGCAAAGGA GTGTATCTAT GTAGAACCGG 2760

ACAGGGAAGG TGACAAGAAA GGTGTGTACT GGTACAACAA TAAGTTATGA GGATGATGGT 2820

GAAGAAATTG TCGACTTTTC TCTTGTCTGT TTGTCTTTTG TTAAAGAAGC TATGCTTCGT 2880

TTTAATAATC TTATTGTCCA TTTTGTTGTG TTATGACATT TTGGCTGCTC ATTATGTTAT 2940

GTGGGAAGTT AGCGTTCAAA TGTTTTGGGT CGG 2973

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION: 1..23
(D) OTHER INFORMATION: /product= "synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GGGCATGTNG ARAANARRTG RTG 23

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: misc_feature
(B) LOCATION: 1..23
(D) OTHER INFORMATION: /product= "synthetic oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GGGCATGTRC TRAANARRTG RTG 23

CLAIMS

1. An isolated nucleic acid fragment comprising a nucleic acid sequence encoding a fatty acid desaturase or a fatty acid desaturase-related enzyme with an amino acid identity of 50% or greater to the polypeptide encoded by SEQ ID NOS:1, 3, 5, 7, 9, 11, or 15.

2. The isolated nucleic acid fragment of Claim 1 wherein the amino acid identity is 60% or greater to the polypeptide encoded by SEQ ID NOS:1, 3, 5, 7, 9, 11, or 15.

3. The isolated nucleic acid fragment of Claim 1 wherein the nucleic acid identity is 90% or greater to SEQ ID NOS:1, 3, 5, 7, 9, 11, or 15.

4. The isolated nucleic acid fragment of Claim 1 wherein said fragment is isolated from an oil-producing plant species.

5. An isolated nucleic acid fragment comprising a nucleic acid sequence encoding a delta-12 fatty acid hydroxylase.

6. A chimeric gene capable of causing altered levels of ricinoleic acid in a transformed plant cell, said chimeric gene comprising a nucleic acid fragment of Claim 5, said fragment operably linked to suitable regulatory sequences.

7. A chimeric gene capable of causing altered levels of fatty acids in a transformed plant cell, said chimeric gene comprising a nucleic acid fragment of any of Claims 1, 2, 3, said fragment operably linked to suitable regulatory sequences.

8. Plants containing a chimeric gene of Claim 6 or Claim 7.

9. Oil obtained from seeds of the plants containing the chimeric genes of Claim 8.

10. A method of producing seed oil containing altered levels of unsaturated fatty acids comprising:

(a) transforming a plant cell of an oil-producing species with a chimeric gene of Claim 5;

(b) growing fertile plants from the transformed plant cells of step (a);

(c) screening progeny seeds from the fertile plants of step (b) for the desired levels of unsaturated fatty acids; and

(d) processing the progeny seed of step (c) to obtain seed oil containing altered levels of unsaturated fatty acids.

11. A method of molecular breeding to obtain altered levels of a fatty acid in seed oil of oil-producing plant species comprising:

(a) making a cross between two varieites of oil-producing species differing in the fatty acid trait;

(b) making a Southern blot of restriction enzyme digested genomic DNA isolated from several progeny plants resulting from the cross of step (a); and

(c) hybridizing the Southern blot with the radiolabelled nucleic acid fragment of Claim 1.

12. A method of RFLP mapping comprising:

(a) making a cross between two varieties of plants;

(b) making a Southern blot of restriction enzyme digested genomic DNA isolated from several progeny plants resulting from the cross of step (a); and

(c) hybridizing the Southern blot with the radiolabelled nucleic acid fragments of Claim 1.

13. A method to isolate nucleic acid fragments encoding fatty acid desaturases and related enzymes, comprising:

(a) comparing SEQ ID NOS:2, 4, 6, 8, 10, or 12 and other fatty acid desaturase polypeptide sequences;

(b) identifying the conserved sequences of 4 or more amino acids obtained in step a;

(c) desiging degenerate oligomers based on the conserved sequences identified in step b; and

(d) using the degenerate oligomers of step c to isolate sequences encoding fatty acid desaturases and desaturase-related enzymes by sequence dependent protocols.

14. An isolated nucleic acid fragment of Claim 1 comprising a nucleic acid sequence encoding a plant microsomal delta-12 fatty acid desaturase.

INTERNATI NAL SEARCH REPORT

Intern al Application No
PCT/US 93/09987

A. CLASSIFICATION OF SUBJECT MATTER
IPC 5 C12N15/53 C12N15/82 C11B1/00 C12Q1/68 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 5 C12N C11B A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JOURNAL OF BIOLOGICAL CHEMISTRY vol. 267, no. 3, 25 January 1992, BALTIMORE, MD US pages 1502 - 1509 MIQUEL, M., ET AL. 'Arabidopsis mutants deficient in polyunsaturated fatty acid synthesis' cited in the application	9
A	see the whole document	1-4

X	THEOR. APPL. GENET. vol. 80, 1990 pages 234 - 240 LEMIEUX, B., ET AL. 'Mutants of Arabidopsis with alterations in seed lipid composition' see the whole document	9

	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
10 January 1994	28 -01- 1994

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Maddox, A

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BIOLOGICAL ABSTRACTS, xol. 95 Philadelphia, PA, US; abstract no. 9224, SMITH, M.A., ET AL. 'Evidence for cytochrome b5 as an electron donor in ricinoleic acid biosynthesis in microsomal preparations from developing castor bean (Ricinus communis L.)' see abstract & BIOCHEM. J. vol. 287, no. 1 , 1992 pages 141 - 144	9
Y	UCLA SYMP. MOL. CELL. BIOL.; NEW SER. vol. 129 , 1990 pages 301 - 309 BROWSE, J., ET AL. 'Strategies for modifying plant lipid composition' see the whole document	1-4,14
Y	SCIENCE vol. 252 , 5 April 1991 , LANCASTER, PA US pages 80 - 87 SOMERVILLE, C., ET AL. 'Plant lipids: Metabolism, mutants, and membranes' see page 82, right column, line 24 - line 27	1-4,14
A	see page 85, right column, last paragraph - page 86, left column	5,6,8,9
A	US,A,5 057 419 (MARTIN) 15 October 1991 see column 10, line 4 - line 24	1-4,7,8
A	WO,A,91 18985 (DU PONT) 12 December 1991 see claim 13	11,12
A	WO,A,91 13972 (CALGENE) 19 September 1991 see page 78, line 1 - line 15	10
A	NATURE vol. 347 , 13 September 1990 , LONDON GB pages 200 - 203 WADA, H., ET AL. 'Enhancement of chilling tolerance of a cyanobacterium by genetic manipulation of fatty acid desaturation' see the whole document	1-4

-/--

1

INTERNATI NAL SEARCH REPORT

Interna   l Application No
PCT/US 93/09987

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BIOLOGICAL ABSTRACTS, xol. 72 1981, Philadelphia, PA, US; abstract no. 41091, MOREAU, R.A., ET AL. 'Recent studies on the enzymatic synthesis of ricinoleic acid by developing castor beans(Ricinus communis)' see abstract & PLANT PHYSIOL. vol. 67, no. 4 , 1981 pages 672 - 676 -----	5

1

INTERNATIONAL SEARCH REPORT

...ormation on patent family members

Interna' 'l Application No
PCT/US 93/09987

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
US-A-5057419	15-10-91	NONE		
WO-A-9118985	12-12-91	AU-A-	7900991	31-12-91
		EP-A-	0537178	21-04-93
WO-A-9113972	19-09-91	EP-A-	0472722	04-03-92

14

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A01H 5/00, 5/10, C12N 15/52, 15/82 — A1

(11) International Publication Number: WO 97/30582

(43) International Publication Date: 28 August 1997 (28.08.97)

(21) International Application Number: PCT/US97/02187

(22) International Filing Date: 6 February 1997 (06.02.97)

(30) Priority Data:
08/597,313 6 February 1996 (06.02.96) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/597,313 (CON)
Filed on 6 February 1996 (06.02.96)

(71) Applicants *(for all designated States except US)*: CARNEGIE INSTITUTION OF WASHINGTON [US/US]; 1530 P. Street, N.W., Washington, DC 20016 (US). MONSANTO COMPANY, INC. [US/US]; 700 Chesterfield Parkway North, St. Louis, MO 63198 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: BROUN, Pierre [FR/US]; 1249 Capuchino, Burlingame, CA 94010 (US). VAN DE LOO, Frank [AU/AU]; 11 Fowles Street, Weston, ACT 2611 (AU). BODDUPALLI, Sekhar, S. [IN/US]; 572 Enchanted Parkway, Manchester, MO 63021 (US). SOMERVILLE, Chris [US/US]; 5 Valley Oak, Portola Valley, CA 94028 (US).

(74) Agents: KOKULIS, Paul, N. et al.; Cushman Darby & Cushman, Intellectual Property Group of Pillsbury Madison & Sutro, 1100 New York Avenue, N.W., Washington, DC 20005 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PRODUCTION OF HYDROXYLATED FATTY ACIDS IN GENETICALLY MODIFIED PLANTS

(57) Abstract

This invention relates to plant fatty acid hydroxylases. Methods to use conserved amino acid or nucleotide sequences to obtain plant fatty acid hydroxylases are described. Also described is the use of cDNA clones encoding a plant hydroxylase to produce a family of hydroxylated fatty acids in transgenic plants. In addition, the use of genes encoding fatty acid hydroxylases or desaturases to alter the level of lipid fatty acid unsaturation in transgenic plants is described.

PRODUCTION OF HYDROXYLATED FATTY ACIDS IN GENETICALLY MODIFIED PLANTS

TECHNICAL FIELD

The present invention concerns the identification of nucleic acid sequences and constructs, and methods related thereto, and the use of these sequences and constructs to produce genetically modified plants for the purpose of altering the fatty acid composition of plant oils, waxes and related compounds.

DEFINITIONS

The subject of this invention is a class of enzymes that introduce a hydroxyl group into several different fatty acids resulting in the production of several different kinds of hydroxylated fatty acids. In particular, these enzymes catalyze hydroxylation of oleic acid to 12-hydroxy oleic acid and icosenoic acid to 14-hydroxy icosenoic acid. Other fatty acids such as palmitoleic and erucic acids may also be substrates. Since it is not possible to refer to the enzyme by reference to a unique substrate or product, the enzyme is referred throughout as kappa hydroxylase to indicate that the enzyme introduces the hydroxyl three carbons distal (i.e., away from the carboxyl carbon of the acyl chain) from a double bond located near the center of the acyl chain.

The following fatty acids are also the subject of this invention: ricinoleic acid, 12-hydroxyoctadec-*cis*-9-enoic acid ($12OH\text{-}18{:}1^{cis\Delta9}$); lesquerolic acid, 14-hydroxy-*cis*-11-icosenoic acid ($14OH\text{-}20{:}1^{cis\Delta11}$); densipolic acid, 12-hydroxyoctadec-*cis*-9,15-dienoic acid ($12OH\text{-}18{:}2^{cis\Delta9,15}$); auricolic acid, 14-hydroxy-*cis*-11,17-icosadienoic acid (14OH-

$20{:}2^{cis\Delta 11,17}$); hydroxyerucic, 16-hydroxydocos-*cis*-13-enoic acid ($16OH\text{-}22{:}1^{cis\Delta 13}$); hydroxypalmitoleic, 12-hydroxyhexadec-*cis*-9-enoic ($12OH\text{-}16{:}1^{cis\Delta 9}$); icosenoic acid ($20{:}1^{cis\Delta 11}$). It will be noted that icosenoic acid is spelled eicosenoic acid in some countries.

BACKGROUND

Extensive surveys of the fatty acid composition of seed oils from different species of higher plants have resulted in the identification of at least 33 structurally distinct monohydroxylated plant fatty acids, and 12 different polyhydroxylated fatty acids that are accumulated by one or more plant species (reviewed by van de Loo et al., 1993). Ricinoleic acid, the principal constituent of the seed oil from the castor plant *Ricinus communis* (L.), is of commercial importance. The present inventors have cloned a gene from this species that encodes a fatty acid hydroxylase, and have used this gene to produce ricinoleic acid in transgenic plants of other species. Some of this scientific evidence has been published by the present inventors (van de Loo et al., 1995).

The use of the castor hydroxylase gene to also produce other hydroxylated fatty acids such as lesquerolic acid, densipolic acid, hydroxypalmitoleic, hydroxyerucic and auricolic acid in transgenic plants is the subject of this invention. In addition, the identification of a gene encoding a homologous hydroxylase from *Lesquerella fendleri*, and the use of this gene to produce these hydroxylated fatty acids in transgenic plants is the subject of this invention.

Castor is a minor oilseed crop. Approximately 50% of the seed weight is oil (triacylglycerol) in which 85-90% of total fatty acids are the hydroxylated fatty acid, ricinoleic acid. Oil pressed or extracted from castor seeds has many industrial uses based upon the properties endowed by the hydroxylated fatty acid. The most important uses are production of paints and varnishes, nylon-type synthetic polymers, resins, lubricants, and cosmetics (Atsmon, 1989).

In addition to oil, the castor seed contains the extremely toxic protein ricin, allergenic proteins, and the alkaloid ricinine. These constituents preclude the use of the untreated seed meal (following oil extraction) as a livestock feed, normally an important economic aspect of oilseed utilization. Furthermore, with the variable nature of castor plants and a lack of investment in breeding, castor has few favorable agronomic characteristics.

For a combination of these reasons, castor is no longer grown in the United States and the development of an alternative domestic source of hydroxylated fatty acids would be attractive. The production of ricinoleic acid, the important constituent of castor oil, in an established oilseed crop through genetic engineering would be a particularly effective means of creating a domestic source.

Because there is no practical source of lesquerolic, densipolic and auricolic acids from plants that are adapted to modern agricultural practices, there is currently no large-scale use of these fatty acids by industry. However, the fatty

acids would have uses similar to those of ricinoleic acid if they could be produced in large quantities at comparable cost to other plant-derived fatty acids (Smith, 1985).

Plant species, such as certain species in the genus *Lesquerella*, that accumulate a high proportion of these fatty acids, have not been domesticated and are not currently considered a practical source of fatty acids (Hirsinger, 1989). This invention represents a useful step toward the eventual production of these and other hydroxylated fatty acids in transgenic plants of agricultural importance.

The taxonomic relationships between plants having similar or identical kinds of unusual fatty acids have been examined (van de Loo et al., 1993). In some cases, particular fatty acids occur mostly or solely in related taxa. In other cases there does not appear to be a direct link between taxonomic relationships and the occurrence of unusual fatty acids. In this respect, ricinoleic acid has now been identified in 12 genera from 10 families (reviewed in van de Loo et al., 1993). Thus, it appears that the ability to synthesize hydroxylated fatty acids has evolved several times independently during the radiation of the angiosperms. This suggested to us that the enzymes which introduce hydroxyl groups into fatty acids arose by minor modifications of a related enzyme.

Indeed, as shown herein, the sequence similarity between Δ12 fatty acid desaturases and the kappa hydroxylase from castor is so high that it is not possible to unambiguously determine whether a particular enzyme is a d saturase or a hydroxylase

on the basis of evidence in the scientific literature. Similarly, a patent application (PCT WO 94/11516) that purports to teach the isolation and use of Δ12 fatty acid desaturases does not teach how to distinguish a hydroxylase from a desaturase. In view of the importance of being able to distinguish between these activities for the purpose of genetic engineering of plant oils, the utility of that application is limited to the several instances where direct experimental evidence (e.g., altered fatty acid composition in transgenic plants) was presented to support the assignment of function. A method for distinguishing between fatty acid desaturases and fatty acid hydroxylases on the basis of amino acid sequence of the enzyme is also a subject of this invention.

A feature of hydroxylated or other unusual fatty acids is that they are generally confined to seed triacylglycerols, being largely excluded from the polar lipids by unknown mechanisms (Battey and Ohlrogge 1989; Prasad et al., 1987). This is particularly intriguing since diacylglycerol is a precursor of both triacylglycerol and polar lipid. With castor microsomes, there is some evidence that the pool of ricinoleoyl-containing polar lipid is minimized by a preference of diacylglycerol acyltransferase for ricinoleate-containing diacylglycerols (Bafor et al., 1991). Analyses of vegetative tissues have generated few reports of unusual fatty acids, other than those occurring in the cuticle. The cuticle contains various hydroxylated fatty acids which are interesterified to produce a high molecular weight polyester which serves a structural role. A small number of other

exceptions exist in which unusual fatty acids are found in tissues other than the seed.

The biosynthesis of ricinoleic acid from oleic acid in the developing endosperm of castor (*Ricinus communis*) has been studied by a variety of methods. Morris (1967) established in double-labeling studies that hydroxylation occurs directly by hydroxyl substitution rather than via an unsaturated-, keto- or epoxy-intermediate. Hydroxylation using oleoyl-CoA as precursor can be demonstrated in crude preparations or microsomes, but activity in microsomes is unstable and variable, and isolation of the microsomes involved a considerable, or sometimes complete loss of activity (Galliard and Stumpf, 1966; Moreau and Stumpf, 1981). Oleic acid can replace oleoyl-CoA as a precursor, but only in the presence of CoA, Mg^{2+} and ATP (Galliard and Stumpf, 1966) indicating that activation to the acyl-CoA is necessary. However, no radioactivity could be detected in ricinoleoyl-CoA (Moreau and Stumpf, 1981). These and more recent observations (Bafor et al., 1991) have been interpreted as evidence that the substrate for the castor oleate hydroxylase is oleic acid esterified to phosphatidylcholine or another phospholipid.

The hydroxylase is sensitive to cyanide and azide, and dialysis against metal chelators reduces activity, which could be restored by addition of $FeSO_4$, suggesting iron involvement in enzyme activity (Galliard and Stumpf, 1966). Ricinoleic acid synthesis requires molecular oxygen (Galliard and Stumpf, 1966; Moreau and Stumpf 1981) and requires NAD(P)H to reduce cytochrome b5 which is thought to be the intermediate electron donor for the

hydroxylase reaction (Smith et al., 1992). Carbon monoxide does not inhibit hydroxylation, indicating that a cytochrome P450 is not involved (Galliard and Stumpf, 1966; Moreau and Stumpf 1981). Data from a study of the substrate specificity of the hydroxylase show that all substrate parameters (i.e., chain length and double bond position with respect to both ends) are important; deviations in these parameters caused reduced activity relative to oleic acid (Howling et al., 1972). The position at which the hydroxyl was introduced, however, was determined by the position of the double bond, always being three carbons distal. Thus, the castor acyl hydroxylase enzyme can produce a family of different hydroxylated fatty acids depending on the availability of substrates. Thus, as a matter of convenience, the enzyme is referred throughout this specification as a kappa hydroxylase (rather than an oleate hydroxylase) to indicate the broad substrate specificity.

The castor kappa hydroxylase has many superficial similarities to the microsomal fatty acyl desaturases (Browse and Somerville, 1991). In particular, plants have a microsomal oleate desaturase active at the Δ12 position. The substrate of this enzyme (Schmidt et al., 1993) and of the hydroxylase (Bafor et al., 1991) appears to be a fatty acid esterified to the *sn-2* position of phosphatidylcholine. When oleate is the substrate, the modification occurs at the same position (Δ12) in the carbon chain, and requires the same cofactors, namely electrons from NADH via cytochrome b_5 and molecular oxygen. Neither enzyme is inhibited

by carbon monoxide (Moreau and Stumpf, 1981), the characteristic inhibitor of cytochrome P450 enzymes.

There do not appear to have been any published biochemical studies of the properties of the hydroxylase enzyme(s) in *Lesquerella*.

Conceptual basis of the invention

The present inventors have described the use of a cDNA clone from castor for the production of ricinoleic acid in transgenic plants. As noted above, biochemical studies had suggested that the castor hydroxylase may not have strict specificity for oleic acid but would also catalyze hydroxylation of other fatty acids such as icosenoic acid ($20{:}1^{cis\Delta 11}$) (Howling et al., 1972). Based on these studies, expression of kappa hydroxylase in transgenic plants of species such as *Brassica napus* and *Arabidopsis thaliana* that accumulate fatty acids such as icosenoic acid ($20{:}1^{cis\Delta 11}$) and erucic acid (13-docosenoic acid; $22{:}1^{cis\Delta 13}$) may cause the accumulation of hydroxylated derivatives of these fatty acids due to the activity of the hydroxylase on these fatty acids. Direct evidence is presented in Example 1 that hydroxlyated derivatives of ricinoleic, lesquerolic, densipolic and auricolic fatty acids are produced in transgenic *Arabidopsis* plants.

Example 2 shows the isolation of a novel kappa hydroxylase gene from *Lesquerella fendleri*.

In view of the high degree of sequence similarity between Δ12 fatty acid desaturases and the castor hydroxylase (van de Loo et al., 1995), the validity of claims (e.g., PCT WO 94/11516) for using a limited set of desaturase or hydroxylase

genes or sequences derived therefrom to identify genes of identical function from other species must be viewed with skepticism. In this application, the present inventors teach a method by which hydroxylase genes can be distinguished from desaturases. The present inventors describe a mechanistic basis for the similar reaction mechanisms of desaturases and hydroxylases. Briefly, the available evidence suggests that fatty acid desaturases have a similar reaction mechanism to the bacterial enzyme methane monooxygenase which catalyses a reaction involving oxygen-atom transfer ($CH_4 \rightarrow CH_3OH$) (van de Loo et al., 1993). The cofactor in the hydroxylase component of methane monooxygenase is termed a μ-oxo bridged diiron cluster (FeOFe). The two iron atoms of the FeOFe cluster are liganded by protein-derived nitrogen or oxygen atoms, and are tightly redox-coupled by the covalently-bridging oxygen atom. The FeOFe cluster accepts two electrons, reducing it to the diferrous state, before oxygen binding. Upon oxygen binding, it is likely that heterolytic cleavage also occurs, leading to a high valent oxoiron reactive species that is stabilized by resonance rearrangements possible within the tightly coupled FeOFe cluster. The stabilized high-valent oxoiron state of methane monooxygenase is capable of proton extraction from methane, followed by oxygen transfer, giving methanol. The FeOFe cofactor has been shown to be directly relevant to plant fatty acid modifications by the demonstration that castor stearoyl-ACP desaturase contains this type of cofactor (Fox et al., 1993).

On the basis of the foregoing considerations, the present inventors suggest that the castor oleate hydroxylase might be a structurally modified fatty acyl desaturase, based upon three arguments. The first argument involves the taxonomic distribution of plants containing ricinoleic acid. Ricinoleic acid has been found in 12 genera of 10 families of higher plants (reviewed in van de Loo et al., 1993). Thus, plants in which ricinoleic acid occurs are found throughout the plant kingdom, yet close relatives of these plants do not contain the unusual fatty acid. This pattern suggests that the ability to synthesize ricinoleic acid has arisen (and been lost) several times independently, and is therefore has recently diverged. In other words, the ability to synthesize ricinoleic acid has evolved rapidly, suggesting that a relatively minor genetic change in the structure of the ancestral enzyme was necessary to accomplish it.

The second argument is that many biochemical properties of castor kappa hydroxylase are similar to those of the microsomal desaturases, as discussed above (e.g., both preferentially act on fatty acids esterified to the sn-2 position of phosphatidylcholine, both use cytochrome b5 as an intermediate electron donor, both are inhibited by cyanide, both require molecular oxygen as a substrate, both are thought to be located in the endoplasmic reticulum).

The third argument stems from the discussion of oxygenase cofactors above, in which it is suggested that the plant membrane bound fatty acid desaturases may have a μ-oxo bridged diiron cluster-type cofactor, and that such cofactors are capable

of catalyzing both fatty acid desaturations and hydroxylations, depending upon the electronic and structural properties of the protein active site.

Taking these three arguments together, the present inventors suggest that kappa hydroxylase of castor endosperm is homologous to the microsomal oleate Δ12 desaturase found in all plants. A number of genes encoding microsomal Δ12 desaturases from various species have recently been cloned (Okuley et al., 1994) and substantial information about the structure of these enzymes is now known (Shanklin et al., 1994). Hence, in the following invention, the present inventors teach how to use structural information to isolate and identify kappa hydroxylase genes. This example teaches the method by which any carbon-monoxide insensitive plant fatty acyl hydroxylase gene can be identified by one skilled in the art.

An unpredicted outcome of our studies on the castor hydroxylase gene in transgenic *Arabidopsis* plants was the discovery that expression of the hydroxylase leads to increased accumulation of oleic acid in seed lipids. Because of the low nucleotide sequence homology between the castor hydroxylase and the Δ12-desaturase (about 67%), it is unlikely that this effect is due to silencing (also called sense-suppression or cosuppression) of the expression of the desaturase gene by the hydroxylase gene. Whatever the basis for the effect, this invention teaches the use of hydroxylase genes to alter the level of fatty acid unsaturation in transgenic plants. This invention also teaches the use of genetically modified hydroxylase and desaturase

genes to achieve directed modification of fatty acid unsaturation levels.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-D show the mass spectra of hydroxy fatty acids standards (Figure 1A, O-TMS-methylricinoleate; Figure 1B, O-TMS-methyl densipoleate; Figure 1C, O-TMS-methyl-lesqueroleate; and Figure 1D, O-TMS-methylauricoleate).

Figure 2 shows the fragmentation pattern of trimethylsilylated methyl esters of hydroxy fatty acids.

Figure 3A shows the gas chromatogram of fatty acids extracted from seeds of wild type *Arabidopsis* plants. Figure 3B shows the gas chromatogram of fatty acids extracted from seeds of transgenic *Arabidopsis* plants containing the fah12 hydroxylase gene. The numbers indicate the following fatty acids: [1] 16:0; [2] 18:0; [3] 18:1cisΔ9; [4] $18:2^{cis\Delta 9,12}$; [5] 20:0; [6] $20:1^{cis\Delta 11}$; [7] $18:3^{cis\Delta 9,12,15}$; [8] $20:2^{cis\Delta 11,14}$; [9] $22:1^{cis\Delta 13}$; [10] ricinoleic acid; [11] densipolic acid; [12] lesquerolic acid; and [13] auricolic acid.

Figures 4A-D show the mass spectra of novel fatty acids found in seeds of transgenic plants. Figure 4A shows the mass spectrum of peak 10 from Figure 3B. Figure 4B shows the mass spectrum of peak 11 from Figure 3B. Figure 4C shows the mass spectrum of peak 12 from Figure 3B. Figure 4D shows the mass spectrum of peak 13 from Figure 3B.

Figure 5 shows the nucleotide sequence of pLesq2 (SEQ ID NO:1).

Figure 6 shows the nucleotide sequence of pLesq3 (SEQ ID NO:2).

Figure 7 shows a Northern blot of total RNA from seeds of *L. fendleri* probed with pLesq2 or pLesq3. S, indicates RNA is from seeds; L, indicates RNA is from leaves.

Figures 8A-B show the nucleotide sequence of genomic clone encoding pLesq-HYD (SEQ ID NO:3), and the deduced amino acid sequence of hydroxylase enzyme encoded by the gene (SEQ ID NO:4).

Figures 9A-B show multiple sequence alignment of deduced amino acid sequences for kappa hydroxylases and microsomal Δ12 desaturases. Abbreviations are: Rcfah12, fah12 hydroxylase gene from *R. communis* (van de Loo et al., 1995); Lffah12, kappa hydroxylase gene from *L. fendleri*; Atfad2, fad2 desaturase from *Arabidopsis thaliana* (Okuley et al., 1994); Gmfad2-1, fad2 desaturase from *Glycine max* (GenBank accession number L43920); Gmfad2-2, fad2 desaturase from *Glycine max* (Genbank accession number L43921); Zmfad2, fad2 desaturase from *Zea mays* (PCT WO 94/11516); Rcfad2, fragment of fad2 desaturase from *R. communis* (PCT WO 94/11516); Bnfad2, fad2 desaturase from *Brassica napus* (PCT WO 94/11516); LFFAH12.AMI, SEQ ID NO:4; FAH12.AMI, SEQ ID NO:5; ATFAD2.AMI, SEQ ID NO:6; BNFAD2.AMI, SEQ ID NO:7; GMFAD2-1.AMI, SEQ ID NO:8; GMFAD2-2.AMI, SEQ ID NO:9; ZMFAD2.AMI, SEQ ID NO:10; and RCFAD2.AMI, SEQ ID NO:11.

Figure 10 shows a Southern blot of genomic DNA from *L. fendleri* probed with pLesq-HYD. E = *Eco*RI, H = *Hin*dIII, X = *Xba*I.

Figure 11 shows a map of binary Ti plasmid pSLJ44024.

Figure 12 shows a map of plasmid pYES2.0

Figure 13 shows part of a gas chromatogram of derivatized fatty acids from yeast cells that contain plasmid pLesqYes in which expression of the hydroxylase gene was induced by addition of galactose to the growth medium. The arrow points to a peak that is not present in uninduced cells. The lower part of the figure is the mass spectrum of the peak indicated by the arrow.

SUMMARY OF THE INVENTION

This invention relates to plant fatty acyl hydroxylases. Methods to use conserved amino acid or nucleotide sequences to obtain plant fatty acyl hydroxylases are described. Also described is the use of cDNA clones encoding a plant hydroxylase to produce a family of hydroxylated fatty acids in transgenic plants.

In a first embodiment, this invention is directed to recombinant DNA constructs which can provide for the transcription, or transcription and translation (expression) of the plant kappa hydroxylase sequence. In particular, constructs which are capable of transcription, or transcription and translation in plant host cells are preferred. Such constructs may contain a variety of regulatory regions including transcriptional initiation regions obtained from genes preferentially expressed in plant seed tissue. In a second aspect, this invention relates to the presence of such constructs in host cells, especially plant host cells which have an expressed plant kappa hydroxylase therein.

In yet another aspect, this invention relates to a method for producing a plant kappa hydroxylase in a host cell or progeny thereof via the expression

of a construct in the cell. Cells containing a plant kappa hydroxylase as a result of the production of the plant kappa hydroxylase encoding sequence are also contemplated herein.

In another embodiment, this invention relates to methods of using a DNA sequence encoding a plant kappa hydroxylase for the modification of the proportion of hydroxylated fatty acids produced within a cell, especially plant cells. Plant cells having such a modified hydroxylated fatty acid composition are also contemplated herein.

In a further aspect of this invention, plant kappa hydroxylase proteins and sequences which are related thereto, including amino acid and nucleic acid sequences, are contemplated. Plant kappa hydroxylase exemplified herein includes a *Lesquerella fendleri* fatty acid hydroxylase. This exemplified fatty acid hydroxylase may be used to obtain other plant fatty acid hydroxylases of this invention.

In a further aspect of this invention, a nucleic acid sequence which directs the seed specific expression of an associated polypeptide coding sequence is described. The use of this nucleic acid sequence or fragments derived therefrom, to obtain seed-specific expression in higher plants of any coding sequence is contemplated herein.

In a further aspect of this invention, the use of genes encoding fatty acyl hydroxylases of this invention are used to alter the amount of fatty acid unsaturation of seed lipids. The present invention further discloses the use of genetically modified hydroxylase and desaturase genes to achieve

directed modification of fatty acid unsaturation levels.

DETAILED DESCRIPTION OF THE INVENTION

A genetically transformed plant of the present invention which accumulates hydroxylated fatty acids can be obtained by expressing the double-stranded DNA molecules described in this application.

A plant fatty acid hydroxylase of this invention includes any sequence of amino acids, such as a protein, polypeptide or peptide fragment, or nucleic acid sequences encoding such polypeptides, obtainable from a plant source which demonstrates the ability to catalyze the production of ricinoleic, lesquerolic, hydroxyerucic (16-hydroxydocos-*cis*-13-enoic acid) or hydroxypalmitoleic (12-hydroxyhexadec-*cis*-9-enoic) from CoA, ACP or lipid-linked monoenoic fatty acid substrates under plant enzyme reactive conditions. By "enzyme reactive conditions" is meant that any necessary conditions are available in an environment (i.e., such factors as temperature, pH, lack of inhibiting substances) which will permit the enzyme to function.

Preferential activity of a plant fatty acid hydroxylase toward a particular fatty acyl substrate is determined upon comparison of hydroxylated fatty acid product amounts obtained per different fatty acyl substrates. For example, by "oleate preferring" is meant that the hydroxylase activity of the enzyme preparation demonstrates a preference for oleate-containing substrates over other substrates. Although the precise substrate of the castor fatty

acid hydroxylase is not known, it is thought to be a monounsaturated fatty acid moiety which is esterified to a phospholipid such as phosphatidylcholine. However, it is also possible that monounsaturated fatty acids esterified to phosphatidylethanolamine, phosphatidic acid or a neutral lipid such as diacylglycerol or a Coenzyme-A thioester may also be substrates.

As noted above, significant activity has been observed in radioactive labelling studies using fatty acyl substrates other than oleate (Howling et al., 1972) indicating that the substrate specificity is for a family of related fatty acyl compounds. Because the castor hydroxylase introduces hydroxy groups three carbons from a double bond, proximal to the methyl carbon of the fatty acid, the enzyme is termed a kappa hydroxylase for convenience. Of particular interest, the present invention discloses that the castor kappa hydroxylase may be used for production of 12-hydroxy-9-octadecenoic acid (ricinoleate), 12-hydroxy-9-hexadecenoic acid, 14-hydroxy-11-eicosenoic acid, 16-hydroxy-13-docosenoic acid, 9-hydroxy-6-octadecenoic acid by expression in plants species which produce the non-hydroxylated precursors. The present invention also discloses production of additionally modified fatty acids such as 12-hydroxy-9,15-octadecadienoic acid that result from desaturation of hydroxylated fatty acids (e.g., 12-hydroxy-9-octadecenoic acid in this example).

The present invention also discloses that future advances in the genetic engineering of plants will lead to production of substrate fatty acids, such as icosenoic acid esters, and palmitoleic acid esters in plants that do not normally accumulate

such fatty acids. The invention described herein may be used in conjunction with such future improvements to produce hydroxylated fatty acids of this invention in any plant species that is amenable to directed genetic modification. Thus, the applicability of this invention is not limited in our conception only to those species that currently accumulate suitable substrates.

As noted above, a plant kappa hydroxylase of this invention will display activity towards various fatty acyl substrates. During biosynthesis of lipids in a plant cell, fatty acids are typically covalently bound to acyl carrier protein (ACP), coenzyme A (CoA) or various cellular lipids. Plant kappa hydroxylases which display preferential activity toward lipid-linked acyl substrate are especially preferred because they are likely to be closely associated with normal pathway of storage lipid synthesis in immature embryos. However, activity toward acyl-CoA substrates or other synthetic substrates, for example, is also contemplated herein.

Other plant kappa hydroxylases are obtainable from the specific exemplified sequences provided herein. Furthermore, it will be apparent that one can obtain natural and synthetic plant kappa hydroxylases including modified amino acid sequences and starting materials for synthetic-protein modeling from the exemplified plant kappa hydroxylase and from plant kappa hydroxylases which are obtained through the use of such exemplified sequences. Modified amino acid sequences include sequences which have been mutated, truncated, elongated or the like, whether such sequences were

partially or wholly synthesized. Sequences which are actually purified from plant preparations or are identical or encode identical proteins thereto, regardless of the method used to obtain the protein or sequence, are equally considered naturally derived.

Thus, one skilled in the art will readily recognize that antibody preparations, nucleic acid probes (DNA and RNA) or the like may be prepared and used to screen and recover "homologous" or "related" kappa hydroxylases from a variety of plant sources. Typically, nucleic acid probes are labeled to allow detection, preferably with radioactivity although enzymes or other methods may also be used. For immunological screening methods, antibody preparations either monoclonal or polyclonal are utilized. Polyclonal antibodies, although less specific, typically are more useful in gene isolation. For detection, the antibody is labeled using radioactivity or any one of a variety of second antibody/enzyme conjugate systems that are commercially available.

Homologous sequences are found when there is an identity of sequence and may be determined upon comparison of sequence information, nucleic acid or amino acid, or through hybridization reactions between a known kappa hydroxylase and a candidate source. Conservative changes, such as Glu/Asp, Val/Ile, Ser/Thr, Arg/Lys and Gln/Asn may also be considered in determining sequence homology. Typically, a lengthy nucleic acid sequence may show as little as 50-60% sequence identity, and more preferably at least about 70% sequence identity, between the target sequence and the given plant

kappa hydroxylase of interest excluding any deletions which may be present, and still be considered related. Amino acid sequences are considered homologous by as little as 25% sequence identity between the two complete mature proteins. (see generally, Doolittle, R.F., OF URFS and ORFS, University Science Books, CA, 1986.)

A genomic or other appropriate library prepared from the candidate plant source of interest may be probed with conserved sequences from the plant kappa hydroxylase to identify homologously related sequences. Use of an entire cDNA or other sequence may be employed if shorter probe sequences are not identified. Positive clones are then analyzed by restriction enzyme digestion and/or sequencing. When a genomic library is used, one or more sequences may be identified providing both the coding region, as well as the transcriptional regulatory elements of the kappa hydroxylase gene from such plant source. Probes can also be considerably shorter than the entire sequence. Oligonucleotides may be used, for example, but should be at least about 10, preferably at least about 15, more preferably at least 20 nucleotides in length. When shorter length regions are used for comparison, a higher degree of sequence identity is required than for longer sequences. Shorter probes are often particularly useful for polymerase chain reactions (PCR), especially when highly conserved sequences can be identified (see Gould et al., 1989 for examples of the use of PCR to isolate homologous genes from taxonomically diverse species).

When longer nucleic acid fragments are employed (>100 bp) as probes, especially when using

complete or large cDNA sequences, one would screen with low stringencies (for example, 40-50°C below the melting temperature of the probe) in order to obtain signal from the target sample with 20-50% deviation, i.e., homologous sequences (Beltz et al., 1983).

In a preferred embodiment, a plant kappa hydroxylase of this invention will have at least 60% overall amino acid sequence similarity with the exemplified plant kappa hydroxylase. In particular, kappa hydroxylases which are obtainable from an amino acid or nucleic acid sequence of a castor or *Lesquerella* kappa hydroxylase are especially preferred. The plant kappa hydroxylases may have preferential activity toward longer or shorter chain fatty acyl substrates. Plant fatty acyl hydroxylases having oleate-12-hydroxylase activity and eicosenoate-14-hydroxylase activity are both considered homologously related proteins because of *in vitro* evidence (Howling et al., 1972), and evidence disclosed herein, that the castor kappa hydroxylase will act on both substrates. Hydroxylated fatty acids may be subject to further enzymatic modification by other enzymes which are normally present or are introduced by genetic engineering methods. For example, 14-hydroxy-11,17-eicosadienoic acid, which is present in some *Lesquerella* species (Smith, 1985), is thought to be produced by desaturation of 14-hydroxy-11-eicosenoic acid.

Again, not only can gene clones and materials derived therefrom be used to identify homologous plant fatty acyl hydroxylases, but the resulting sequences obtained therefrom may also provide a

further method to obtain plant fatty acyl hydroxylases from other plant sources. In particular, PCR may be a useful technique to obtain related plant fatty acyl hydroxylases from sequence data provided herein. One skilled in the art will be able to design oligonucleotide probes based upon sequence comparisons or regions of typically highly conserved sequence. Of special interest are polymerase chain reaction primers based on the conserved regions of amino acid sequence between the castor kappa hydroxylase and the *L. fendleri* hydroxylase (SEQ ID NO:4). Details relating to the design and methods for a PCR reaction using these probes are described more fully in the examples.

It should also be noted that the fatty acyl hydroxylases of a variety of sources can be used to investigate fatty acid hydroxylation events in a wide variety of plant and *in vivo* applications. Because all plants synthesize fatty acids via a common metabolic pathway, the study and/or application of one plant fatty acid hydroxylase to a heterologous plant host may be readily achieved in a variety of species.

Once the nucleic acid sequence is obtained, the transcription, or transcription and translation (expression), of the plant fatty acyl hydroxylases in a host cell is desired to produce a ready source of the enzyme and/or modify the composition of fatty acids found therein in the form of free fatty acids, esters (particularly esterified to glycerolipids or as components of wax esters), estolides, or ethers. Other useful applications may be found when the host cell is a plant host cell, *in vitro* and *in vivo*. For example, by increasing the amount of an kappa

hydroxylase available to the plant, an increased percentage of ricinoleate or lesqueroleate (14-hydroxy-11-eicosenoic acid) may be provided.

Kappa Hydroxylase

By this invention, a mechanism for the biosynthesis of ricinoleic acid in plants is demonstrated. Namely, that a specific plant kappa hydroxylase having preferential activity toward fatty acyl substrates is involved in the accumulation of hydroxylated fatty acids in at least some plant species. The use of the terms ricinoleate or ricinoleic acid (or lesqueroleate or lesquerolic acid, densipoleate etc.) is intended to include the free acids, the ACP and CoA esters, the salts of these acids, the glycerolipid esters (particularly the triacylglycerol esters), the wax esters, the estolides and the ether derivatives of these acids.

The determination that plant fatty acyl hydroxylases are active in the *in vivo* production of hydroxylated fatty acids suggests several possibilities for plant enzyme sources. In fact, hydroxylated fatty acids are found in some natural plant species in abundance. For example, three hydroxy fatty acids related to ricinoleate occur in major amounts in seed oils from various *Lesquerella* species. Of particular interest, lesquerolic acid is a 20 carbon homolog of ricinoleate with two additional carbons at the carboxyl end of the chain (Smith, 1985). Other natural plant sources of hydroxylated fatty acids include but are not limited to seeds of the *Linum* genus, seeds of *Wrightia* species, *Lycopodium* species, *Strophanthus* species,

Convolvulaces species, *Calendula* species and many others (van de Loo et al., 1993).

Plants having significant presence of ricinoleate or lesqueroleate or desaturated other or modified derivatives of these fatty acids are preferred candidates to obtain naturally-derived kappa hydroxylases. For example, *Lesquerella densipila* contains a diunsaturated 18 carbon fatty acid with a hydroxyl group (van de Loo et al., 1993) that is thought to be produced by an enzyme that is closely related to the castor kappa hydroxylase, according to the theory on which this invention is based. In addition, a comparison between kappa hydroxylases and between plant fatty acyl hydroxylases which introduce hydroxyl groups at positions other than the 12-carbon of oleate or the 14-carbon of lesqueroleate or on substrates other than oleic acid and icosenoic acid may yield insights for gene identification, protein modeling or other modifications as discussed above.

Especially of interest are fatty acyl hydroxylases which demonstrate activity toward fatty acyl substrates other than oleate, or which introduce the hydroxyl group at a location other than the C12 carbon. As described above, other plant sources may also provide sources for these enzymes through the use of protein purification, nucleic acid probes, antibody preparations, protein modeling, or sequence comparisons, for example, and of special interest are the respective amino acid and nucleic acid sequences corresponding to such plant fatty acyl hydroxylases. Also, as previously described, once a nucleic acid sequence is obtained for the given plant hydroxylase, further plant

sequences may be compared and/or probed to obtain homologously related DNA sequences thereto and so on.

Genetic Engineering Applications

As is well known in the art, once a cDNA clone encoding a plant kappa hydroxylase is obtained, it may be used to obtain its corresponding genomic nucleic acid sequences thereto.

The nucleic acid sequences which encode plant kappa hydroxylases may be used in various constructs, for example, as probes to obtain further sequences from the same or other species. Alternatively, these sequences may be used in conjunction with appropriate regulatory sequences to increase levels of the respective hydroxylase of interest in a host cell for the production of hydroxylated fatty acids or study of the enzyme *in vitro* or *in vivo* or to decrease or increase levels of the respective hydroxylase of interest for some applications when the host cell is a plant entity, including plant cells, plant parts (including but not limited to seeds, cuttings or tissues) and plants.

A nucleic acid sequence encoding a plant kappa hydroxylase of this invention may include genomic, cDNA or mRNA sequence. By "encoding" is meant that the sequence corresponds to a particular amino acid sequence either in a sense or anti-sense orientation. By "recombinant" is meant that the sequence contains a genetically engineered modification through manipulation via mutagenesis, restriction enzymes, or the like. A cDNA sequence may or may not encode pre-processing sequences, such

as transit or signal peptide sequences. Transit or signal peptide sequences facilitate the delivery of the protein to a given organelle and are frequently cleaved from the polypeptide upon entry into the organelle, releasing the "mature" sequence. The use of the precursor DNA sequence is preferred in plant cell expression cassettes.

Furthermore, as discussed above the complete genomic sequence of the plant kappa hydroxylase may be obtained by the screening of a genomic library with a probe, such as a cDNA probe, and isolating those sequences which regulate expression in seed tissue.

Once the desired plant kappa hydroxylase nucleic acid sequence is obtained, it may be manipulated in a variety of ways. Where the sequence involves non-coding flanking regions, the flanking regions may be subjected to resection, mutagenesis, etc. Thus, transitions, transversions, deletions, and insertions may be performed on the naturally occurring sequence. In addition, all or part of the sequence may be synthesized. In the structural gene, one or more codons may be modified to provide for a modified amino acid sequence, or one or more codon mutations may be introduced to provide for a convenient restriction site or other purpose involved with construction or expression. The structural gene may be further modified by employing synthetic adapters, linkers to introduce one or more convenient restriction sites, or the like.

The nucleic acid or amino acid sequences encoding a plant kappa hydroxylase of this invention may be combined with other non-native, or "heterologous", sequences in a variety of ways. By

"heterologous" sequences is meant any sequence which is not naturally found joined to the plant kappa hydroxylase, including, for example, combination of nucleic acid sequences from the same plant which are not naturally found joined together.

The DNA sequence encoding a plant kappa hydroxylase of this invention may be employed in conjunction with all or part of the gene sequences normally associated with the kappa hydroxylase. In its component parts, a DNA sequence encoding kappa hydroxylase is combined in a DNA construct having, in the 5' to 3' direction of transcription, a transcription initiation control region capable of promoting transcription and/or translation in a host cell, the DNA sequence encoding plant kappa hydroxylase and a transcription and/or translation termination region.

Potential host cells include both prokaryotic and eukaryotic cells. A host cell may be unicellular or found in a multicellular differentiated or undifferentiated organism depending upon the intended use. Cells of this invention may be distinguished by having a plant kappa hydroxylase foreign to the wild-type cell present therein, for example, by having a recombinant nucleic acid construct encoding a plant kappa hydroxylase therein.

Depending upon the host, the regulatory regions will vary, including regions from viral, plasmid or chromosomal genes, or the like. For expression in prokaryotic or eukaryotic microorganisms, particularly unicellular hosts, a wide variety of constitutive or regulatable promoters may be employed. Expression in a

microorganism can provide a ready source of the plant enzyme. Among transcriptional initiation regions which have been described are regions from bacterial and yeast hosts, such as *E. coli*, *B. subtilis*, *Saccharomyces cerevisiae*, including genes such as beta-galactosidase, T7 polymerase, trpE or the like.

For the most part, the constructs will involve regulatory regions functional in plants which provide for modified production of plant kappa hydroxylase with resulting modification of the fatty acid composition. The open reading frame, coding for the plant kappa hydroxylase or functional fragment thereof will be joined at its 5' end to a transcription initiation regulatory region. Numerous transcription initiation regions are available which provide for a wide variety of constitutive or regulatable, e.g., inducible, transcription of the structural gene functions.

Among transcriptional initiation regions used for plants are such regions associated with the structural genes such as for nopaline and mannopine synthases, or with napin, soybean β-conglycinin, oleosin, 12S storage protein, the cauliflower mosaic virus 35S promoters or the like. The transcription/ translation initiation regions corresponding to such structural genes are found immediately 5' upstream to the respective start codons.

In embodiments wherein the expression of the kappa hydroxylase protein is desired in a plant host, the use of all or part of the complete plant kappa hydroxylase gene is desired. If a different promoter is desired, such as a promoter native to the plant host of interest or a modified promoter,

i.e., having transcription initiation regions derived from one gene source and translation initiation regions derived from a different gene source or enhanced promoters, such as double 35S CaMV promoters, the sequences may be joined together using standard techniques.

For such applications when 5' upstream non-coding regions are obtained from other genes regulated during seed maturation, those preferentially expressed in plant embryo tissue, such as transcription initiation control regions from the *B. napus* napin gene, or the *Arabidopsis* 12S storage protein, or soybean β-conglycinin (Bray et al., 1987) are desired. Transcription initiation regions which are preferentially expressed in seed tissue, i.e., which are undetectable in other plant parts, are considered desirable for fatty acid modifications in order to minimize any disruptive or adverse effects of the gene product.

Regulatory transcript termination regions may be provided in DNA constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding the plant kappa hydroxylase or a convenient transcription termination region derived from a different gene source, for example, the transcript termination region which is naturally associated with the transcript initiation region. Where the transcript termination region is from a different gene source, it will contain at least about 0.5 kb, preferably about 1-3 kb of sequence 3' to the structural gene from which the termination region is derived.

Plant expression or transcription constructs having a plant kappa hydroxylase as the DNA sequence

of interest for increased or decreased expression thereof may be employed with a wide variety of plant life, particularly, plant life involved in the production of vegetable oils for edible and industrial uses. Most especially preferred are temperate oilseed crops. Plants of interest include, but are not limited to rapeseed (canola and high erucic acid varieties), *Crambe, Brassica juncea, Brassica nigra*, meadowfoam, flax, sunflower, safflower, cotton, *Cuphea*, soybean, peanut, coconut and oil palms and corn. An important criterion in the selection of suitable plants for the introduction on the kappa hydroxylase is the presence in the host plant of a suitable substrate for the hydroxylase. Thus, for example, production of ricinoleic acid will be best accomplished in plants that normally have high levels of oleic acid in seed lipids. Similarly, production of lesquerolic acid will best be accomplished in plants that have high levels of icosenoic acid in seed lipids.

Depending on the method for introducing the recombinant constructs into the host cell, other DNA sequences may be required. Importantly, this invention is applicable to dicotyledons and monocotyledons species alike and will be readily applicable to new and/or improved transformation and regulation techniques. The method of transformation is not critical to the current invention; various methods of plant transformation are currently available. As newer methods are available to transform crops, they may be directly applied hereunder. For example, many plant species naturally susceptible to *Agrobacterium* infection may be successfully transformed via tripartite or binary

vector methods of *Agrobacterium* mediated transformation. In addition, techniques of microinjection, DNA particle bombardment, electroporation have been developed which allow for the transformation of various monocot and dicot plant species.

In developing the DNA construct, the various components of the construct or fragments thereof will normally be inserted into a convenient cloning vector which is capable of replication in a bacterial host, e.g., *E. coli*. Numerous vectors exist that have been described in the literature. After each cloning, the plasmid may be isolated and subjected to further manipulation, such as restriction, insertion of new fragments, ligation, deletion, insertion, resection, etc., so as to tailor the components of the desired sequence. Once the construct has been completed, it may then be transferred to an appropriate vector for further manipulation in accordance with the manner of transformation of the host cell.

Normally, included with the DNA construct will be a structural gene having the necessary regulatory regions for expression in a host and providing for selection of transformant cells. The gene may provide for resistance to a cytotoxic agent, e.g., antibiotic, heavy metal, toxin, etc., complementation providing prototropy to an auxotrophic host, viral immunity or the like. Depending upon the number of different host species the expression construct or components thereof are introduced, one or more markers may be employed, where different conditions for selection are used for the different hosts.

It is noted that the degeneracy of the DNA code provides that some codon substitutions are permissible of DNA sequences without any corresponding modification of the amino acid sequence.

As mentioned above, the manner in which the DNA construct is introduced into the plant host is not critical to this invention. Any method which provides for efficient transformation may be employed. Various methods for plant cell transformation include the use of Ti- or Ri-plasmids, microinjection, electroporation, infiltration, imbibition, DNA particle bombardment, liposome fusion, DNA bombardment or the like. In many instances, it will be desirable to have the construct bordered on one or both sides of the T-DNA, particularly having the left and right borders, more particularly the right border. This is particularly useful when the construct uses *A. tumefaciens* or *A. rhizogenes* as a mode for transformation, although the T-DNA borders may find use with other modes of transformation.

Where *Agrobacterium* is used for plant cell transformation, a vector may be used which may be introduced into the *Agrobacterium* host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the *Agrobacterium* host. The Ti- or Ri-plasmid containing the T-DNA for recombination may be armed (capable of causing gall formation) or disarmed (incapable of causing gall), the latter being permissible, so long as the *vir* genes are present in the transformed *Agrobacterium* host. The armed plasmid can give a mixture of normal plant cells and gall.

In some instances where *Agrobacterium* is used as the vehicle for transforming plant cells, the expression construct bordered by the T-DNA border(s) will be inserted into a broad host spectrum vector, there being broad host spectrum vectors described in the literature. Commonly used is pRK2 or derivatives thereof. See, for example, Ditta et al. (1980), which is incorporated herein by reference. Included with the expression construct and the T-DNA will be one or more markers, which allow for selection of transformed *Agrobacterium* and transformed plant cells. A number of markers have been developed for use with plant cells, such as resistance to kanamycin, the aminoglycoside G418, hygromycin, or the like. The particular marker employed is not essential to this invention, one or another marker being preferred depending on the particular host and the manner of construction.

For transformation of plant cells using *Agrobacterium*, explants may be combined and incubated with the transformed *Agrobacterium* for sufficient time for transformation, the bacteria killed, and the plant cells cultured in an appropriate selective medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of vegetable oils.

Using Hydroxylase Genes to Alter the Activity of Fatty Acid Desaturases

A widely acknowledged goal of current efforts to improve the nutritional quality of edible plant oils, or to facilitate industrial applications of plant oils, is to alter the level of desaturation of plant storage lipids (Topfer et al., 1995). In particular, in many crop species it is considered desirable to reduce the level of polyunsaturation of storage lipids and to increase the level of oleic acid. The precise amount of the various fatty acids in a particular plant oil varies with the intended application. Thus, it is desirable to have a robust method that will permit genetic manipulation of the level of unsaturation to any desired level.

Substantial progress has recently been made in the isolation of genes encoding plant fatty acid desaturases (reviewed in Topfer et al., 1995). These genes have been introduced into various plant species and used to alter the level of fatty acid unsaturation in one of three ways. First, the genes can be placed under transcriptional control of a strong promoter so that the amount of the corresponding enzyme is increased. In some cases this leads to an increase in the amount of the fatty acid that is the product of the reaction catalyzed by the enzyme. For example, Arondel et al. (1992) increased the amount of linolenic acid (18:3) in tissues of transgenic *Arabidopsis* plants by placing the endoplasmic reticulum-localized fad3 gene under transcriptional control of the strong constitutive cauliflower mosaic virus 35S promoter.

A second method of using cloned genes to alter the level of fatty acid unsaturation is to

cause transcription of all or part of a gene in transgenic tissues so that the transcripts have an antisense orientation relative to the normal mode of transcription. This has been used by a number of laboratories to reduce the level of expression of one or more desaturase genes that have significant nucleotide sequence homology to the gene used in the construction of the antisense gene (reviewed in Topfer et al.). For instance, antisense repression of the oleate Δ12-desaturase in transgenic rapeseed resulted in a strong increase in oleic acid content (cf., Topfer et al., 1995).

A third method for using cloned genes to alter fatty acid desaturation is to exploit the phenomenon of cosuppression or "gene-silencing" (Matzke et al., 1995). Although the mechanisms responsible for gene silencing are not known in any detail, it has frequently been observed that in transgenic plants, expression of an introduced gene leads to inactivation of homologous endogenous genes.

For example, high-level sense expression of the *Arabidopsis* fad8 gene, which encodes a chloroplast-localized Δ15-desaturase, in transgenic *Arabidopsis* plants caused suppression of the endogenous copy of the fad8 gene and the homologous fad7 gene (which encodes an isozyme of the fad8 gene) (Gibson et al., 1994). The fad7 and fad8 genes are only 76% identical at the nucleotide level. At the time of publication, this example represented the most divergent pair of plant genes for which cosuppression had been observed.

In view of previous evidence concerning the relatively high level of nucleotide sequence

homology required to obtain cosuppression, it is not obvious to one skilled in the art that sense expression in transgenic plants of the castor fatty acyl hydroxylase of this invention would significantly alter the amount of unsaturation of storage lipids.

However, the present inventors establish that fatty acyl hydroxylase genes can be used for this purpose as taught in Example 4 of this specification. Of particular importance, this invention teaches the use of fatty acyl hydroxylase genes to increase the proportion of oleic acid in transgenic plant tissues. The mechanism by which expression of the gene exerts this effect is not known but may be due to one of several possibilities which are elaborated upon in Example 4.

The invention now being generally described, it will be more readily understood by reference to the following examples which are included for purposes of illustration only and are not intended to limit the present invention.

EXAMPLES

In the experimental disclosure which follows, all temperatures are given in degrees centigrade (°C), weights are given in grams (g), milligram (mg) or micrograms (μg), concentrations are given as molar (M), millimolar (mM) or micromolar (μM) and all volumes are given in liters (l), microliters (μl) or milliliters (ml), unless otherwise indicated.

EXAMPLE 1 - PRODUCTION OF NOVEL HYDROXYLATED FATTY ACIDS IN *ARABIDOPSIS THALIANA*

Overview

The kappa hydroxylase encoded by the fah12 gene from castor was used to produce ricinoleic acid, lesquerolic acid, densipolic acid and auricolic acid in transgenic *Arabidopsis* plants.

Production of transgenic plants

A variety of methods have been developed to insert a DNA sequence of interest into the genome of a plant host to obtain the transcription and translation of the sequence to effect phenotypic changes. The following methods represent only one of many equivalent means of producing transgenic plants and causing expression of the hydroxylase gene.

Arabidopsis plants were transformed, by *Agrobacterium*-mediated transformation, with the kappa hydroxylase encoded by the castor fah12 gene on binary Ti plasmid pB6. This plasmid has also been used to transform *Nicotiana tabacum* for the production of ricinoleic acid.

Inoculums of *Agrobacterium tumefaciens* strain GV3101 containing binary Ti plasmid pB6 were plated on L-broth plates containing 50 µg/ml kanamycin and incubated for 2 days at 30°C. Single colonies were used to inoculate large liquid cultures (L-broth medium with 50 mg/l rifampicin, 110 mg/l gentamycin and 200 mg/l kanamycin) to be used for the transformation of *Arabidopsis* plants.

Arabidopsis plants were transformed by the *in planta* transformation procedure essentially as described by Bechtold et al. (1993). Cells of *A. tumefaciens* GV3101(pB6) were harvested from liquid

cultures by centrifugation, then resuspended in infiltration medium at OD_{600} = 0.8. Infiltration medium was Murashige and Skoog macro and micronutrient medium (Sigma Chemical Co., St. Louis, MO) containing 10 mg/l 6-benzylaminopurine and 5% glucose. Batches of 12-15 plants were grown for 3 to 4 weeks in natural light at a mean daily temperature of approximately 25°C in 3.5 inch pots containing soil. The intact plants were immersed in the bacterial suspension then transferred to a vacuum chamber and placed under 600 mm of vacuum produced by a laboratory vacuum pump until tissues appeared uniformly water-soaked (approximately 10 min). The plants were grown at 25°C under continuous light (100 μmol m^{-2} s^{-1} irradiation in the 400 to 700 nm range) for four weeks. The seeds obtained from all the plants in a pot were harvested as one batch. The seeds were sterilized by sequential treatment for 2 min with ethanol followed by 10 min in a mixture of household bleach (Chlorox), water and Tween-80 (50%, 50%, 0.05%) then rinsed thoroughly with sterile water. The seeds were plated at high density (2000 to 4000 per plate) onto agar-solidified medium in 100 mm petri plates containing 1/2 X Murashige and Skoog salts medium enriched with B5 vitamins (Sigma Chemical Co., St. Louis, MO) and containing kanamycin at 50 mg/l. After incubation for 48 h at 4°C to stimulate germination, seedlings were grown for a period of seven days until transformants were clearly identifiable as healthy green seedlings against a background of chlorotic kanamycin-sensitive seedlings. The transformants were transferred to soil for two weeks before leaf tissue

could be used for DNA and lipid analysis. More than 20 transformants were obtained.

DNA was extracted from young leaves from transformants to verify the presence of an intact fah12 gene. The presence of the transgene in a number of the putative transgenic lines was verified by using the polymerase chain reaction to amplify the insert from pB6. The primers used were HF2 = GCTCTTTTGTGCGCTCATTC (SEQ ID NO:12) and HR1 = CGGTACCAGAAAACGCCTTG (SEQ ID NO:13), which were designed to allow the amplification of a 700 bp fragment. Approximately 100 ng of genomic DNA was added to a solution containing 25 pmol of each primer, 1.5 U Taq polymerase (Boehringer Manheim), 200 uM of dNTPs, 50 mM KCl, 10 mM Tris.Cl (pH 9), 0.1% (v/v) Triton X-100, 1.5 mM $MgCl_2$, 3% (v/v) formamide, to a final volume of 50 μl. Amplifications conditions were: 4 min denaturation step at 94°C, followed by 30 cycles of 92°C for 1 min, 55°C for 1 min, 72°C for 2 min. A final extension step closed the program at 72°C for 5 min. Transformants could be positively identified after visualization of a characteristic 1 kb amplified fragment on an ethidium bromide stained agarose gel. All transgenic lines tested gave a PCR product of a size consistent with the expected genotype, confirming that the lines were, indeed, transgenic. All further experiments were done with three representative transgenic lines of the wild type designated as 1-3, 4D, 7-4 and one transgenic line of the fad2 mutant line JB12. The transgenic JB12 line was included in order to test whether the increased accumulation of oleic acid in this mutant

would have an effect on the amount of ricinoleic acid that accumulated in the transgenic plants.

Analysis of transgenic plants

Leaves and seeds from fah12 transgenic *Arabidopsis* plants were analyzed for the presence of hydroxylated fatty acids using gas chromatography. Lipids were extracted from 100-200 mg leaf tissue or 50 seeds. Fatty acid methyl esters (FAMES) were prepared by placing tissue in 1.5 ml of 1.0 M methanolic HCl (Supelco Co.) in a 13 x 100 mm glass screw-cap tube capped with a teflon-lined cap and heated to 80°C for 2 hours. Upon cooling, 1 ml petroleum ether was added and the FAMES removed by aspirating off the ether phase which was then dried under a nitrogen stream in a glass tube. One hundred μl of N,O-bis(Trimethylsilyl) trifluoroacetamide (BSTFA; Pierce Chemical Co) and 200 μl acetonitrile was added to derivatize the hydroxyl groups. The reaction was carried out at 70°C for 15 min. The products were dried under nitrogen, redissolved in 100 μl chloroform and transferred to a gas chromatograph vial. Two μl of each sample were analyzed on a SP2340 fused silica capillary column (30 m, 0.75 mm ID, 0.20 mm film, Supelco), using a Hewlett-Packard 5890 II series Gas Chromatograph. The samples were not split, the temperature program was 195°C for 18 min, increased to 230°C at 25°C/min, held at 230°C for 5 min then down to 195°C at 25°C/min., and flame ionization detectors were used.

The chromatographic elution time of methyl esters and O-TMS derivatives of ricinoleic acid, lesquerolic acid and auricolic acid was established

by GC-MS of lipid samples from seeds of *L. fendleri* and comparison to published chromatograms of fatty acids from this species (Carlson et al., 1990). A O-TMS-methyl-ricinoleate standard was prepared from ricinoleic acid obtained from Sigma Chemical Co (St, Louis, MO). O-TMS-methyl-lesqueroleate and O-TMS-methyl-auricoleate standards were prepared from triacylglycerols purified from seeds of *L. fendleri*. The mass spectrum of O-TMS-methyl-ricinoleate, O-TMS-methyl-densipoleate, O-TMS-methyl-lesqueroleate, and O-TMS-methyl-auricoleate are shown in Figures 1A-D, respectively. The structures of the characteristic ions produced during mass spectrometry of these derivatives are shown in Figure 2.

Lipid extracted from transgenic tissues were analyzed by gas chromatography and mass spectrometry for the presence of hydroxylated fatty acids. As a matter of reference, the average fatty acid composition of leaves in *Arabidopsis* wild type and fad2 mutant lines was reported by Miquel and Browse (1992). Gas chromatograms of methylated and silylated fatty acids from seeds of wild type and a fah12 transgenic wild type plant are shown in Figures 3A and 3B, respectively. The profiles are very similar except for the presence of three small but distinct peaks at 14.3, 15.9 and 18.9 minutes. A very small peak at 20.15 min was also evident. The elution time of the peaks at 14.3 and 18.9 min corresponded precisely to that of comparably prepared ricinoleic and lesquerolic standards, respectively. No significant differences were observed in lipid extracts from leaves or roots of

the wild type and the fah12 transgenic wild type lines (Table 1).

Thus, in spite of the fact that the fah12 gene is expressed throughout the plant, effects on fatty acid composition was observed only in seed tissue. The present inventors have made a similar observation for transgenic fah12 tobacco.

Table 1. Fatty acid composition of lipids from transgenic and wild type *Arabidopsis*. The values are the means obtained from analysis of samples from three independent transgenic lines, or three independent samples of wild type and fad2 lines.

TABLE 1

Fatty acid	Seed				Leaf		Root	
	WT	FAH12 WT	FAH12 fad2	JB12	WT	FAH12 WT	WT	FAH12 WT
16:0	8.5	8.2	6.4	6.1	16.5	17.5	23.9	24.9
16:3	0	0	0	0	10.1	9.8	0	0
18:0	3.2	3.5	2.9	3.5	1.3	1.2	2.0	1.9
18:1	15.4	26.3	43.4	47.8	2.4	3.4	5.4	3.2
18:2	27.0	21.4	10.2	7.2	15.1	14.0	32.2	29.4
18:3	22.0	16.6	-	9.7	36.7	36.0	26.7	30.6
20:1	14.0	14.3	-	13.1	0	0	0	0

TABLE 1 (continued)

Fatty acid	Seed				Leaf		Root	
	WT	FAH12 WT	FAH12 fad2	JB12	WT	FAH12 WT	WT	FAH12 WT
18:1-OH	0	0.4	0.3	0	0	0	0	0
18:2-OH	0	0.4	0.3	0	0	0	0	0
20:1-OH	0	0.2	0.1	0	0	0	0	0
20:2-OH	0	0.1	0.1	0	0	0	0	0

In order to confirm that the observed new peaks in the transgenic lines corresponded to derivatives of ricinoleic, lesquerolic, densipolic and auricolic acids, mass spectrometry was used. The fatty acid derivatives were resolved by gas chromatography as described above except that a Hewlett-Packard 5971 series mass selective detector was used in place of the flame ionization detector used in the previous experiment. The spectra of the four new peaks in Figure 3B (peak numbers 10, 11, 12 and 13) are shown in Figures 4A-D, respectively. Comparison of the spectrum obtained for the standards with that obtained for the four peaks from the transgenic lines confirms the identity of the four new peaks. On the basis of the three characteristic peaks at M/Z 187, 270 and 299, peak 10 is unambiguously identified as O-TMS-methylricinoleate. On the basis of the three characteristic peaks at M/Z 185, 270 and 299, peak 11 is unambiguously identified as O-TMS-methyldensipoleate. On the basis of the three characteristic peaks at M/Z 187, 298 and 327, peak 12 is unambiguously identified as O-TMS-methyllesqueroleate. On the basis of the three characteristic peaks at M/Z 185, 298 and 327, peak 13 is unambiguously identified as O-TMS-methylauricoleate.

These results unequivocally demonstrate the identity of the fah12 cDNA as encoding a hydroxylase that hydroxylates both oleic acid to produce ricinoleic acid and also hydroxylates icosenoic acid to produce lesquerolic acid. These results also provide additional evidence that the hydroxylase can be functionally expressed in a heterologous plant

species in such a way that the enzyme is catalytically functional. These results also demonstrate that expression of this hydroxylase gene leads to accumulation of ricinoleic, lesquerolic, densipolic and auricolic acids in a plant species that does not normally accumulate hydroxylated fatty acids in extractable lipids.

The present inventors expected to find lesquerolic acid in the transgenic plants based on the biochemical evidence suggesting broad substrate specificity of the kappa hydroxylase. By contrast, the accumulation of densipolic and auricolic acids was less predictable. Since *Arabidopsis* does not normally contain significant quantities of the non-hydroxylated precursors of these fatty acids which could serve as substrates for the hydroxylase, it appears that one or more of the three n-3 fatty acid desaturases known in *Arabidopsis* (e.g., fad3, fad7, fad8; reviewed in Gibson et al., 1995) are capable of desaturating the hydroxylated compounds at the n-3 position. That is, densipolic acid is produced by the action of an n-3 desaturase on ricinoleic acid. Auricolic acid is produced by the action of an n-3 desaturase on lesquerolic acid. Because it is located in the endoplasmic reticulum, the fad3 desaturase is almost certainly responsible. This can be tested in the future by producing fah12-containing transgenic plants of the fad3-deficient mutant of *Arabidopsis* (similar experiments can be done with fad7 and fad8). It is also formally possible that the enzymes that normally elongate $18{:}1^{cis\Delta9}$ to $20{:}1^{cis\Delta11}$ may elongate $12OH\text{-}18{:}1^{cis\Delta9}$ to $14OH\text{-}20{:}1^{cis\Delta11}$, and $12OH\text{-}18{:}2^{cis\Delta9,15}$ to $14OH\text{-}20{:}2^{cis\Delta11,17}$.

The amount of the various fatty acids in seed, leaf and root lipids of the control and transgenic plants is also presented in Table 1. Although the amount of hydroxylated fatty acids produced in this example is less than desired for production of ricinoleate and other hydroxylated fatty acids from plants, numerous improvements may be envisioned that will increase the level of accumulation of hydroxylated fatty acids in plants that express the fah12 or related hydroxylase genes. Improvements in the level and tissue specificity of expression of the hydroxylase gene are envisioned. Methods to accomplish this by the use of strong, seed-specific promoters such as the *B. napus* napin promoter will be obvious to one skilled in the art. Additional improvements are envisioned that involve modification of the enzymes which cleave hydroxylated fatty acids from phosphatidylcholine, reduction in the activities of enzymes which degrade hydroxylated fatty acids and replacement of acyltransferases which transfer hydroxylated fatty acids to the sn-1, sn-2 and sn-3 positions of glycerolipids. Although genes for these enzymes have not been described in the scientific literature, their utility in improving the level of production of hydroxylated fatty acids can be readily appreciated based on the results of biochemical investigations of ricinoleate synthesis.

Although *Arabidopsis* is not an economically important plant species, it is widely accepted by plant biologists as a model for higher plants. Therefore, the inclusion of this example is intended to demonstrate the general utility of the invention described here to the modification of oil

composition in higher plants. One advantage of studying the expression of this novel gene in *Arabidopsis* is the existence in this system of a large body of knowledge on lipid metabolism, as well as the availability of a collection of mutants which can be used to provide useful information on the biochemistry of fatty acid hydroxylation in plant species. Another advantage is the ease of transposing any of the information obtained on metabolism of ricinoleate in *Arabidopsis* to closely related species such as the crop plants *Brassica napus, Brassica juncea* or *Crambe abyssinica* in order to mass produce ricinoleate, lesqueroleate or other hydroxylated fatty acids for industrial use. The kappa hydroxylase is useful for the production of ricinoleate or lesqueroleate in any plant species that accumulates significant levels of the precursors, oleic acid and icosenoic acid. Of particular interest are genetically modified varieties that accumulate high levels of oleic acid. Such varieties are currently available for sunflower and canola. Production of lesquerolic acid and related hydroxy fatty acids can be achieved in species that accumulate high levels of icosenoic acid or other long chain monoenoic acids. Such plants may in the future be produced by genetic engineering of plants that do not normally make such precursors. Thus, the use of the kappa hydroxylase will be of general utility.

EXAMPLE 2. ISOLATION OF *LESQUERELLA* KAPPA HYDROXYLASE GENOMIC CLONE

Overview

Regions of nucleotide sequence that were conserved in both the castor kappa hydroxylase and the *Arabidopsis* fad2 Δ12 fatty acid desaturase were used to design oligonucleotide primers. These were used with genomic DNA from *Lesquerella fendleri* to amplify fragments of several homologous genes. These amplified fragments were then used as hybridization probes to identify full length genomic clones from a genomic library of *L. fendleri*.

Hydroxylated fatty acids are specific to the seed tissue of *Lesquerella* sp., and are not found to any appreciable extent in vegetative tissues. One of the two genes identified by this method was expressed in both leaves and developing seeds and is therefore thought to correspond to the Δ12 fatty acid desaturase. The other gene was expressed at high levels in developing seeds but was not expressed or was expressed at very low levels in leaves and is the kappa hydroxylase from this species. The identity of the gene as a fatty acyl hydroxylase was established by functional expression of the gene in yeast.

The identity of this gene will also be established by introducing the gene into transgenic *Arabidopsis* plants and showing that it causes the accumulation of ricinoleic acid, lesquerolic acid, densipolic acid and auricolic acid in seed lipids.

The various steps involved in this process are described in detail below. Unless otherwise indicated, routine methods for manipulating nucleic

acids, bacteria and phage were as described by Sambrook et al. (1989).

Isolation of a fragment of the *Lesquerella* kappa hydroxylase gene

Oligonucleotide primers for the amplification of the *L. fendleri* kappa hydroxylase were designed by choosing regions of high deduced amino acid sequence homology between the castor kappa hydroxylase and the *Arabidopsis* Δ12 desaturase (fad2). Because most amino acids are encoded by several different codons, these oligonucleotides were designed to encode all possible codons that could encode the corresponding amino acids.

The sequence of these mixed oligonucleotides was Oligo 1: TAYWSNCAYMGNMGNCAYCA (SEQ ID NO:14) and Oligo 2: RTGRTGNGCNACRTGNGTRTC (SEQ ID NO:15) where Y = C+T, W = A+T, S = G+C, N = A+G+C+T, M = A+C, and R = A+G.

These oligonucleotides were used to amplify a fragment of DNA from *L. fendleri* genomic DNA by the polymerase chain reaction (PCR) using the following conditions: Approximately 100 ng of genomic DNA was added to a solution containing 25 pmol of each primer, 1.5 U Taq polymerase (Boehringer Manheim), 200 uM of dNTPs, 50 mM KCl, 10 mM Tris.Cl (pH 9), 0.1% (v/v) Triton X-100, 1.5 mM $MgCl_2$, 3% (v/v) formamide, to a final volume of 50 μl. Amplifications conditions were: 4 min denaturation step at 94°C, followed by 30 cycles of 92°C for 1 min, 55°C for 1 min, 72°C for 2 min. A final extension step closed the program at 72°C for 5 min.

PCR products of approximately 540 bp were observed following electrophoretic separation of the

products of the PCR reaction in agarose gels. Two of these fragments were cloned into pBluescript (Stratagene) to give rise to plasmids pLesq2 and pLesq3. The sequence of the inserts in these two plasmids was determined by the chain termination method. The sequence of the insert in pLesq2 is presented as Figure 5 (SEQ ID NO:1) and the sequence of the insert in pLesq3 is presented as Figure 6 (SEQ ID NO:2). The high degree of sequence identity between the two clones indicated that they were both potential candidates to be either a Δ12 desaturase or a kappa hydroxylase.

<u>Northern analysis</u>

In *L. fendleri*, hydroxylated fatty acids are found in large amounts in seed oils but are not found in appreciable amounts in leaves. An important criterion in discriminating between a fatty acyl desaturase and kappa hydroxylase is that the kappa hydroxylase gene is expected to be expressed more highly in tissues which have high level of hydroxylated fatty acids than in other tissues. In contrast, all plant tissues should contain mRNA for an ω6 fatty acyl desaturase since diunsaturated fatty acids are found in the lipids of all tissues in most or all plants.

Therefore, it was of great interest to determine whether the gene corresponding to pLesq2 was also expressed only in seeds, or is also expressed in other tissues. This question was addressed by testing for hybridization of pLesq2 to RNA purified from developing seeds and from leaves.

Total RNA was purified from developing seeds and young leaves of *L. fendleri* using an Rneasy RNA

extraction kit (Qiagen), according to the manufacturer's instructions. RNA concentrations were quantified by UV spectrophotometry at λ=260 and 280 nm. In order to ensure even loading of the gel to be used for Northern blotting, RNA concentrations were further adjusted after recording fluorescence under UV light of RNA samples stained with ethidium bromide and run on a test denaturing gel.

Total RNA prepared as described above from leaves and developing seeds was electrophoresed through an agarose gel containing formaldehyde (Iba et al., 1993). An equal quantity (10 μg) of RNA was loaded in both lanes, and RNA standards (0.16-1.77 kb ladder, Gibco-BRL) were loaded in a third lane. Following electrophoresis, RNA was transferred from the gel to a nylon membrane (Hybond N+, Amersham) and fixed to the filter by exposure to UV light.

A ^{32}P-labelled probe was prepared from insert DNA of clone pLesq2 by random priming and hybridized to the membrane overnight at 52°C, after it had been prehybridized for 2 h. The prehybridization solution contained 5X SSC, 10X Denhardt's solution, 0.1% SDS, 0.1M KPO_4 pH 6.8, 100 μg/ml salmon sperm DNA. The hybridization solution had the same basic composition, but no SDS, and it contained 10% dextran sulfate and 30% formamide. The blot was washed once in 2X SSC, 0.5% SDS at 65°C then in 1X SSC at the same temperature.

Brief (30 min) exposure of the blot to X-ray film revealed that the probe pLesq2 hybridized to a single band only in the seed RNA lane (Figure 7). The blot was re-probed with the insert from pLesq3 gene, which gave bands of similar intensity in the seed and leaf lanes (Figure 7).

These results show that the gene corresponding to the clone pLesq2 is highly and specifically expressed in seed of *L. fendleri*. In conjunction with knowledge of the nucleotide and deduced amino acid sequence, strong seed-specific expression of the gene corresponding to the insert in pLesq2 is a convincing indicator of the role of the enzyme in synthesis of hydroxylated fatty acids in the seed oil.

Characterization of a genomic clone of the kappa hydroxylase

Genomic DNA was prepared from young leaves of *L. fendleri* as described by Murray and Thompson (1980). A *Sau3A*I-partial digest genomic library constructed in the vector λDashII (Stratagene, 11011 North Torrey Pines Road, La Jolla CA 92037) was prepared by partially digesting 500 μg of DNA, size-selecting the DNA on a sucrose gradient (Sambrook et al., 1989), and ligating the DNA (12 kb average size) to the *BamH*I-digested arms of λDashII. The entire ligation was packaged according to the manufacturer's conditions and plated on *E. coli* strain XL1-Blue MRA-P2 (Stratagene). This yielded $5x10^5$ primary recombinant clones. The library was then amplified according to the manufacturer's conditions. A fraction of the genomic library was plated on *E. coli* XL1-Blue and resulting plaques (150,000) were lifted to charged nylon membranes (Hybond N+, Amersham), according to the manufacturer's conditions. DNA was crosslinked to the filters under UV in a Stratalinker (Stratagene).

Several clones carrying genomic sequences corresponding to the *L. fendleri* hydroxylase were

isolated by probing the membranes with the insert from pLesq2 that was PCR-amplified with internal primers and labelled with ^{32}P by random priming. The filters were prehybridized for 2 hours at 65°C in 7% SDS, 1mM EDTA, 0.25 M Na_2HPO_4 (pH 7.2), 1% BSA and hybridized to the probe for 16 hours in the same solution. The filters were sequentially washed at 65°C in solutions containing 2 X SSC, 1 X SSC, 0.5 X SSC in addition to 0.1 % SDS. A 2.6 kb *Xba*I fragment containing the complete coding sequence for the kappa hydroxylase and approximately 1 kb of the 5' upstream region was subcloned into the corresponding site of pBluescript KS to produce plasmid pLesq-Hyd and the sequence determined completely using an automatic sequencer by the dideoxy chain termination method. Sequence data was analyzed using the program DNASIS (Hitachi Company).

The sequence of the insert in clone pLesq-Hyd is shown in Figures 8A-B. The sequence entails 1855 bp of contiguous DNA sequence (SEQ ID NO:3). The clone encodes a 401 bp 5' untranslated region (i.e., nucleotides preceding the first ATG codon), an 1152 bp open reading frame, and a 302 bp 3' untranslated region. The open reading frame encodes a 384 amino acid protein with a predicted molecular weight of 44,370 (SEQ ID NO:4). The amino terminus lacks features of a typical signal peptide (von Heijne, 1985).

The exact translation-initiation methionine has not been experimentally determined, but on the basis of deduced amino acid sequence homology to the castor kappa hydroxylase (noted below) is thought to be the methionine encoded by the first ATG codon at nucleotide 402.

Comparison of the pLesq-Hyd deduced amino acid sequence with sequences of membrane-bound desaturases and the castor hydroxylase (Figures 9A-B) indicates that pLesq-Hyd is homologous to these genes. This figure shows an alignment of the *L. fendleri* hydroxylase (SEQ ID NO:4) with the castor hydroxylase (van de Loo et al., 1995), the *Arabidopsis fad2* cDNA which encodes an endoplasmic reticulum-localized Δ12 desaturase (called fad2) (Okuley et al., 1994), two soybean fad2 desaturase clones, a *Brassica napus* fad2 clone, a *Zea mays* fad2 clone and partial sequence of a *R. communis* fad2 clone.

The high degree of sequence homology indicates that the gene products are of similar function. For instance, the overall homology between the *Lesquerella* hydroxylase and the *Arabidopsis* fad2 desaturase was 92.2% similarity and 84.8% identity and the two sequences differed in length by only one amino acid.

Southern hybridization

Southern analysis was used to examine the copy number of the genes in the *L. fendleri* genome corresponding to the clone pLesq-Hyd. Genomic DNA (5 μg) was digested with *Eco*RI, *Hin*dIII and *Xba*I and separated on a 0.9% agarose gel. DNA was alkali-blotted to a charged nylon membrane (Hybond N+, Amersham), according to the manufacturer's protocol. The blot was prehybridized for 2 hours at 65°C in 7% SDS, 1mM EDTA, 0.25 M Na_2HPO_4 (pH 7.2), 1% BSA and hybridized to the probe for 16 hours in the same solution with pLesq-Hyd insert PCR-amplified with internal primers and labelled with ^{32}P by random

priming. The filters were sequentially washed at 65°C in solutions containing 2 X SSC, 1 X SSC, 0.5 X SSC in addition to 0.1 % SDS, then exposed to X-ray film.

The probe hybridized with a single band in each digest of *L. fendleri* DNA (Figure 10), indicating that the gene from which pLesq-Hyd was transcribed is present in a single copy in the *L. fendleri* genome.

<u>Expression of pLesq-Hyd in Transgenic Plants</u>

There are a wide variety of plant promoter sequences which may be used to cause tissue-specific expression of cloned genes in transgenic plants. For instance, the napin promoter and the acyl carrier protein promoters have previously been used in the modification of seed oil composition by expression of an antisense form of a desaturase (Knutson et al., 1992). Similarly, the promoter for the β-subunit of soybean β-conglycinin has been shown to be highly active and to result in tissue-specific expression in transgenic plants of species other than soybean (Bray et al., 1987). Thus, other promoters which lead to seed-specific expression may also be employed for the production of modified seed oil composition. Such modifications of the invention described here will be obvious to one skilled in the art.

Constructs for expression of *L. fendleri* kappa hydroxylase in plant cells are prepared as follows: A 13 kb *Sal*I fragment containing the pLesq-Hyg gene was ligated into the *Xho*I site of binary Ti plasmid vector pSLJ44026 (Jones et al., 1992) (Figure 11) to produce plasmid pTi-Hyd and

transformed into *Agrobacterium tumefaciens* strains GV3101 by electroporation. Strain GV3101 (Koncz and Schell, 1986) contains a disarmed Ti plasmid. Cells for electroporation were prepared as follows. GV3101 was grown in LB medium with reduced NaCl (5 g/l). A 250 ml culture was grown to OD_{600} = 0.6, then centrifuged at 4000 rpm (Sorvall GS-A rotor) for 15 min. The supernatant was aspirated immediately from the loose pellet, which was gently resuspended in 500 ml ice-cold water. The cells were centrifuged as before, resuspended in 30 ml ice-cold water, transferred to a 30 ml tube and centrifuged at 5000 rpm (Sorvall SS-34 rotor) for 5 min. This was repeated three times, resuspending the cells consecutively in 30 ml ice-cold water, 30 ml ice-cold 10% glycerol, and finally in 0.75 ml ice-cold 10% glycerol. These cells were aliquoted, frozen in liquid nitrogen, and stored at -80°C.

Electroporations employed a Biorad Gene Pulser instrument using cold 2 mm-gap cuvettes containing 40 μl cells and 1 μl of DNA in water, at a voltage of 2.5 KV, and 200 Ohms resistance. The electroporated cells were diluted with 1 ml SOC medium (Sambrook et al., 1989, page A2) and incubated at 28°C for 2-4 h before plating on medium containing kanamycin (50 mg/l).

Arabidopsis thaliana can be transformed with the *Agrobacterium* cells containing pTi-Hyd as described in Example 1 above. Similarly, the presence of hydroxylated fatty acids in the transgeneic *Arabidopsis* plants can be demonstrated by the methods described in Example 1 above.

Constitutive expression of the *L. fendleri* hydroxylase in transgenic plants

A 1.5 kb *Eco*RI fragment from pLesq-Hyg comprising the entire coding region of the hydroxylase was gel purified, then cloned into the corresponding site of pBluescript KS (Stratagene). Plasmid DNA from a number of recombinant clones was then restricted with *Pst*I, which should cut only once in the insert and once in the vector polylinker sequence. Release of a 920 bp fragment with *Pst*I indicated the right orientation of the insert for further manipulations. DNA from one such clone was further restricted with *Sal*I, the 5′ overhangs filled-in with the Klenow fragment of DNA polymerase I, then cut with *Sac*I. The insert fragment was gel purified, and cloned between the *Sma*I and *Sac*I sites of pBI121 (Clontech) behind the cauliflower mosaic virus 35S promoter. After checking that the sequence of the junction between insert and vector DNA was appropriate, plasmid DNA from a recombinant clone was used to transform *A. tumefaciens* (GV3101). Kanamycin resistant colonies were then used for *in planta* transformation of *A. thaliana* as previously described.

DNA was extracted from kanamycin resistant seedlings and used to PCR-amplify selected fragments from the hydroxylase using nested primers. When fragments of the expected size could be amplified, corresponding plants were grown in the greenhouse or on agar plates, and fatty acids extracted from fully expanded leaves, roots and dry seeds. GC-MS analysis was then performed as previously described to characterize the different fatty acid species and

detect accumulation of hydroxy fatty acids in transgenic tissues.

<u>Expression of the *Lesquerella* hydroxylase in yeast</u>

In order to demonstrate that the cloned *L. fendleri* gene encoded a kappa hydroxylase, the gene was expressed in yeast cells under transcriptional control of an inducible promoter and the yeast cells were examined for the presence of hydroxylated fatty acids by GC-MS.

In a first step, a lambda genomic clone containing the *L. fendleri* hydroxylase gene was cut with *Eco*RI, and a resulting 1400 bp fragment containing the coding sequence of the hydroxylase gene was subcloned in the *Eco*RI site of the pBluescript KS vector (Stratagene). This subclone, pLesqcod, contains the coding region of the *Lesquerella* hydroxylase plus some additional 3' sequence.

In a second step, pLesqcod was cut with *Hin*dIII and *Xba*I, and the insert fragment was cloned into the corresponding sites of the yeast expression vector pYes2 (Invitrogen; Figure 12). This subclone, pLesqYes, contains the *L. fendleri* hydroxylase in the sense orientation relative to the 3' side of the Gal1 promoter. This promoter is inducible by the addition of galactose to the growth medium, and is repressed upon addition of glucose. In addition, the vector carries origins of replication allowing the propagation of pLesqYes in both yeast and *E. coli*.

<u>Transformation of *S. cerevisiae* host strain CGY2557</u>

Yeast strain CGY2557 (*MATα*, GAL⁺, *ura3-52, leu2-3, trp1, ade2-1, lys2-1, his5, can1-100*) was

grown overnight at 28°C in YPD liquid medium (10 g yeast extract, 20 g bacto-peptone, 20 g dextrose per liter), and an aliquot of the culture was inoculated into 100 ml fresh YPD medium and grown until the OD_{600} of the culture was 1. Cells were then collected by centrifugation and resuspended in about 200μl of supernatant. 40μl aliquots of the cell suspension were then mixed with 1-2μg DNA and electroporated in 2 mm-gap cuvettes using a Biorad Gene Pulser instrument set at 600 V, 200 Ω, 25 μF, 160μl YPD was added and the cells were plated on selective medium containing glucose. Selective medium consisted of 6.7 g yeast nitrogen base (Difco), 0.4 g casamino acids (Difco), 0.02 g adenine sulfate, 0.03 g L-leucine, 0.02 g L-tryptophan, 0.03 g L-lysine-HCl, 0.03 g L-histidine-HCl , 2% glucose, water to 1 liter. Plates were solidified using 1.5% Difco Bacto-agar. Transformant colonies appeared after 3 to 4 days incubation at 28°C.

Expression of the *L. fendleri* hydroxylase in yeast

Independent transformant colonies from the previous experiment were used to inoculate 5 ml of selective medium containing either 2% glucose (gene repressed) or 2% galactose (gene induced) as the sole carbon source. Independent colonies of CGY2557 transformed with pYES2 containing no insert were used as controls.

After 2 days of growth at 28°C, an aliquot of the cultures was used to inoculate 5 ml of fresh selective medium. The new culture was placed at 16°C and grown for 9 days.

Fatty acid analysis of yeast expressing the *L. fendleri* hydroxylase

Cells from 2.5 ml of culture were pelleted at 1800g, and the supernatant was aspirated as completely as possible. Pellets were then dispersed in 1 ml of 1 N methanolic HCl (Supelco, Bellafonte, PA). Transmethylation and derivatization of hydroxy fatty acids were performed as described above. After drying under nitrogen, samples were redissolved in 50µl chloroform before being analyzed by GC-MS. Samples were injected into an SP2330 fused-silica capillary column (30 m x 0.25 mm ID, 0.25µm film thickness, Supelco). The temperature profile was 100 - 160°C, 25°C/min, 160 - 230°C, 10°C/min, 230°C, 3 min, 230-100°C, 25°C/min. Flow rate was 0.9 ml/min. Fatty acids were analyzed using a Hewlett-Packard 5971 series Msdetector.

Gas chromatograms of derivatized fatty acid methyl esters from induced cultures of yeast containing pLesqYes contained a novel peak that eluted at 7.6 min (Figure 13). O-TMS methyl ricinoleate eluted at exactly the same position on control chromatograms. This peak was not present in cultures lacking pLesqYes or in cultures containing pLesqYes grown on glucose (repressing conditions) rather than galactose (inducing conditions). Mass spectrometry of the peak (Figure 13) revealed that the peak has the same spectrum as O-TMS methyl ricinoleate. Thus, on the basis of chromatographic retention time and mass spectrum, it was concluded that the peak corresponded to O-TMS methyl ricinoleate. The presence of ricinoleate in the transgenic yeast cultures confirms the identity of the gene as a kappa hydroxylase of this invention.

EXAMPLE 3. OBTAINING OTHER PLANT FATTY ACYL HYDROXYLASES

The castor fah12 sequence could be used to identify other kappa hydroxylases by methods such as PCR and heterologous hybridization. However, because of the high degree of sequence similarity between Δ12 desaturases and kappa hydroxylases, the prior art does not teach how to distinguish between the two kinds of enzymes without a functional test such as demonstrating activity in transgenic plants or another suitable host (e.g., transgenic microbial or animal cells). The identification of the *L. fendleri* hydroxylase provided for the development of criteria by which a hydroxylase and a desaturase may be distinguished solely on the basis of deduced amino acid sequence information.

Figures 9A-B show a sequence alignment of the castor and *L. fendleri* hydroxylase sequences with the castor hydroxylase sequence and all publicly available sequences for all plant microsomal Δ12 fatty acid desaturases. Of the 384 amino acid residues in the castor hydroxylase sequence, more than 95% are identical to the corresponding residue in at least one of the desaturase sequences. Therefore, none of these residues are responsible for the catalytic differences between the hydroxylase and the desaturases. Of the remaining 16 residues in the castor hydroxylase and 14 residues in the *Lesquerella* hydroxylase, all but seven represent instances where the hydroxylase sequence has a conservative substitution compared with one or more of the desaturase sequences, or there is wide variability in the amino acid at that position in the various desaturases. By conservative, it is

meant that the following amino acids are functionally equivalent: Ser/Thr, Ile/Leu/Val/Met, Asp/Glu. Thus, these structural differences also cannot account for the catalytic differences between the desaturases and hydroxylases. This leaves just seven amino acid residues where both the castor hydroxylase and the *Lesquerella* hydroxylase differ from all of the known desaturases and where all of the known microsomal Δ12 desaturases have the identical amino acid residue. These residues occur at positions 69, 111, 155, 226, 304, 331 and 333 of the alignment in Figure 9. Therefore, these seven sites distinguish hydroxylases from desaturases. Based on this analysis, the present inventors believe that any enzyme with greater than 60% sequence identity to one of the enzymes listed in Figure 9 can be classified as a hydroxylase if it differs from the sequence of the desaturases at these seven positions. Because of slight differences in the number of residues in a particular protein, the numbering may vary from protein to protein but the intent of the number system will be evident if the protein in question is aligned with the castor hydroxylase using the numbering system shown herein. Thus, in conjunction with the methods for using the *Lesquerella* hydroxylase gene to isolate homologous genes, the structural criterion disclosed here teaches how to isolate and identify plant kappa hydroxylase genes for the purpose of genetically modifying fatty acid composition as disclosed herein.

EXAMPLE 4 - USING HYDROXYLASES TO ALTER THE LEVEL OF FATTY ACID UNSATURATION

Evidence that kappa hydroxylases of this invention can be used to alter the level of fatty acid unsaturation was obtained from the analysis of transgenic plants that expressed the castor hydroxylase under control of the cauliflower mosaic virus promoter. The construction of the plasmids and the production of transgenic *Arabidopsis* plants was described in Example 1 (above). The fatty acid composition of seed lipids from wild type and six transgenic lines (1-2/a, 1-2/b, 1-3/b, 4F, 7E, 7F) is shown in Table 2.

Table 2. Fatty acid composition of lipids from *Arabidopsis* seeds. The asterisk (*) indicates that for some of these samples, the 18:3 and 20:1 peaks overlapped on the gas chromatograph and, therefore, the total amount of these two fatty acids is reported.

TABLE 2

Fatty acid	WT	1-2/a	1-2/b	1-3/b	4F	7E	7F
16:0	10.3	8.6	9.5	8.4	8.1	8.4	9
18:0	3.5	3.8	3.9	3.3	3.5	3.8	4.2
18:1	14.7	33	34.5	25.5	27.5	30.5	28.5
18:2	32.4	16.9	21	27.5	21.1	20.1	19.8
18:3	13.8	-	14.4	14.8	-	-	-
20:0	1.3	1.6	1	1.1	2.4	1.8	2
20:1	22.5	-	14.1	17.5	-	-	-
18:3 20:1*	-	31.2	-	-	32.1	30.8	30.6
Ricinoleic	0	0.6	0	0.1	0.2	0.7	0.9
Densipolic	0	0.6	0	0.1	0.2	0.5	0.6
Lesquerolic	0	0.2	0	0	0.2	0.2	0.6
Auricolic	0	0.1	0	0	0	0.1	0.1

The results in Table 2 show that expression of the castor hydroxylase in transgenic *Arabidopsis* plants caused a substantial increase in the amount of oleic acid (18:1) in the seed lipids and an approximately corresponding decrease in the amount of linoleic acid (18:2). The average amount of oleic acid in the six transgenic lines was 29.9% versus 14.7% in the wild type.

The precise mechanism by which expression of the castor hydroxylase gene causes increased accumulation of oleic acid is not known. However, an understanding of the mechanism is not required in order to exploit this invention for the directed alteration of plant lipid fatty acid composition. Furthermore, it will be recognized by one skilled in the art that many improvements of this invention may be envisioned. Of particular interest will be the use of other promoters which have high levels of seed-specific expression.

Since hydroxylated fatty acids were not detected in the seed lipids of transgenic line 1-2b, it seems likely that it is not the presence of hydroxylated fatty acids per se that causes the effect of the castor hydroxylase gene on desaturase activity. Protein-protein interaction between the hydroxylase and the Δ12-oleate desaturase or another protein may be required for the overall reaction (e.g., cytochrome b5) or for the regulation of desaturase activity. For example, interaction between the hydroxylase and this other protein may suppress the activity of the desaturase. In particular, the quaternary structure of the membrane-bound desaturases has not been established. It is possible that these enzymes are active as

dimers or as multimeric complexes containing more than two subunits. Thus, if dimers or multimers form between the desaturase and the hydroxylase, the presence of the hydroxylase in the complex may disrupt the activity of the desaturase.

Transgenic plants may be produced in which the hydroxylase enzyme has been rendered inactive by the elimination of one or more of the histidine residues that have been proposed to bind iron molecules required for catalysis. Several of these histidine residues have been shown to be essential for desaturase activity by site directed mutagenesis (Shanklin et al., 1994). Codons encoding histidine residues in the castor hydroxylase gene will be changed to alanine residues as described by Shanklin et al. (1994). The modified genes will be introduced into transgenic plants of *Arabidopsis*, and possibly other species such as tobacco, by the methods described in Example 1 of this application.

In order to examine the effect on all tissues, the strong constitutive cauliflower mosaic virus promoter may be used to cause transcription of the modified genes. However, it will be recognized that in order to specifically examine the effect of expression of the mutant gene on seed lipids, a seed-specific promoter such as the *B. napus* napin promoter may be used. An expected outcome is that expression of the inactive hydroxylase protein in transgenic plants will inhibit the activity of the endoplasmic reticulum-localized Δ12-desaturase. Maximum inhibition will be obtained by expressing high levels of the mutant protein.

In a further embodiment of this invention, mutations that inactivate other hydroxylases, such

as the *Lesquerella* hydroxylase of this invention, may also be useful for decreasing the amount of endoplasmic reticulum-localized Δ12-desaturase activity in the same way as the castor gene. In a further embodiment of this invention, similar mutations of desaturase genes may also be used to inactivate endogenous desaturases. Thus, expression of catalytically inactive fad2 gene from *Arabidopsis* in transgenic *Arabidopsis* may inhibit the activity of the endogenous fad2 gene product.

Similarly, expression of the catalytically inactive forms of Δ12-desaturase from *Arabidopsis* or other plants in transgenic soybean, rapeseed, *Crambe*, *Brassica juncea*, canola, flax, sunflower, safflower, cotton, cuphea, soybean, peanut, coconut, oil palm or corn may lead to inactivation of endogenous Δ12-desaturase activity in these plants. In a further embodiment of this invention, expression of catalytically inactive forms of other desaturases such as the Δ15-desaturases may lead to inactivation of the corresponding desaturases.

An example of a class of mutants useful in the present invention are "dominant negative" mutants that block the function of a gene at the protein level (Herskowitz, 1987). A cloned gene is altered so that it encodes a mutant product capable of inhibiting the wild type gene product in a cell, thus causing the cell to be deficient in the function of that gene product. Inhibitory variants of a wild type product can be designed because proteins have multiple functional domains that can be mutated independently, e.g., oligomerization, substrate binding, catalysis, membrane association domains or the like. In general, dominant negative

proteins retain an intact, functional subset of the domains of the parent, wild type protein, but have the complement of that subset either missing or altered so as to be nonfunctional.

Whatever the precise basis for the inhibitory effect of the castor hydroxylase on desaturation, because the castor hydroxylase has very low nucleotide sequence homology (i.e., about 67%) to the *Arabidopsis* fad2 gene (encoding the endoplasmic reticulum-localized Δ12-desaturase), the inhibitory effect of this gene, which is provisionally called "protein-mediated inhibition" ("protibition"), may have broad utility because it does not depend on a high degree of nucleotide sequence homology between the transgene and the endogenous target gene. In particular, the castor hydroxylase may be used to inhibit the endoplasmic reticulum-localized Δ12-desaturase activity of all higher plants. Of particular relevance are those species used for oil production. These include but are not limited to rapeseed, *Crambe*, *Brassica juncea*, canola, flax, sunflower, safflower, cotton, cuphea, soybean, peanut, coconut, oil palm and corn.

CONCLUDING REMARKS

By the above examples, demonstration of critical factors in the production of novel hydroxylated fatty acids by expression of a kappa hydroxylase gene from castor in transgenic plants is described. In addition, a complete cDNA sequence of the *Lesquerella fendleri* kappa hydroxylase is also provided. A full sequence of the castor hydroxylase is also given with various constructs for use in host cells. Through this invention, one can obtain

the amino acid and nucleic acid sequences which encode plant fatty acyl hydroxylases from a variety of sources and for a variety of applications. Also revealed is a novel method by which the level of fatty acid desaturation can be altered in a directed way through the use of genetically altered hydroxylase or desaturase genes.

All publications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications are herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

REFERENCES

Arondel, V., B. Lemieux, I. Hwang, S. Gibson, H. Goodman, C.R. Somerville. Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. Science 258, 1353-1355 (1992).

Atsmon, D. (1989) Castor, in *Oil Crops of the World*, Robbelen, G., Downey, K.R., and Ashri, A., Eds., McGraw-Hill, New York, pp. 438-447.

Bechtold, N., Ellis, J. and Pelletier, G. (1993) In Planta Agrobacterium mediated gene transfer by infiltration of adult Arabidopsis thaliana plants. C.R. Acad. Sci. Paris 316, 1194-1199.

Beltz, G.A., Jacobs, K.A., Eickbuch, T.H., Cherbas, P.T., Kafatos, F.C. (1983) Isolation of multigene families and determination of homologies by filter hybridization methods. Methods in Enzymology 100, 266-285.

Bray, E.A., Naito, S., Pan, N.S., Anderson, E., Dube, P., Beachy, R.N. (1987) Expression of the β-subunit of β-conglycinin in seeds of transgenic plants. Planta 172, 364-370.

Carlson, K.D., Chaudhry, A., Bagby, M.O (1990) Analysis of oil and meal from lesquerella fendleri seed. J. Am. Oil Chem. Soc. 67, 438-442.

Ditta, G., Stanfield, S., Corbin, D., Helinski, D.R. (1980) Broad host range DNA cloning system for gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti*. Proc. Natl. Acad. Sci. USA 77, 7347-7351.

Gibson, S., Arondel, V., Iba, K., Somerville, C.R. (1994) Temperature Regulated Expression of a Gene Encoding a Chloroplast omega-3 Desaturase from *Arabidopsis thaliana*. Plant Physiol. 106, 1615-1621.

Gould, S.J., Subramani, S., Scheffler, I.E. (1989) Use of the DNA polymerase chain reaction for

homology probing. Proc. Natl. Acad. Sci. USA 86, 1934-1938.

Herskowitz, I. (1987) Functional inactivation of genes by dominant negative mutations. Nature 329, 219-222.

Hirsinger, F. (1989) New oil crops, in *Oil Crops of the World,* Robbelen, G., Downey, K.R., and Ashri, A., Eds., McGraw-Hill, New York, pp. 518-533.

Howling, D., Morris, L.J., Gurr, M.I., James, A.T. (1972) The specificity of fatty acid desaturases and hydroxylases. The dehydrogenation and hydroxylation of monoenoic acids, Biochim. Biophys. Acta 260, 10.

Huyuh, T.V., Young, R.A., Davis, R.W. (1985) Constructing and screening cDNA libraries in λgt10 and λgt11. In DNA Cloning, Vol. 1: A Practical Approach, (ed) D.M. Glover. IRL Press, Washington DC pp 49-77.

Iba, K., Gibson, S., Nishiuchi, T., Fuse, T., Nishimura, M., Arondel, V., Hugly, S., and Somerville, C. (1993) A gene encoding a chloroplast omega-3 fatty acid desaturase complements alterations in fatty acid desaturation and chloroplast copy number of the *fad7* mutant of *Arabidopsis thaliana*. J. Biol. Chem. 268, 24099-24105.

Jones, J.D.G., Shlumukov, L., Carland, F., English, J., Scofield, S., Bishop, G.J., Harrison, K. (1992) Effective vectors for transformation, expression of

heterologous genes, and assaying transposon excision in transgenic plants. Transgenic Res. 1, 285-297.

Knutson, D.S., Thompson, G.A., Radke, S.E., Johnson, W.B., Knauf, V.C., Kridl, J.C. (1992) Proc. Natl. Acad. Sci. USA 89, 2624-2628.

Koncz, C., Schell, J. (1986) The promoter of T_L-DNA gene 5 controls the tissue-specific expression of chimeric genes carried by a novel type of *Agrobacterium* binary vector. Mol. Gen. Genet. 204, 383-396.

Matzke, M., Matzke, A.J.M. (1995) How and why do plants inactivate homologous (Trans)genes? Plant Physiol. 107, 679-685.

Miquel, M. Browse, J. (1992) Arabidopsis mutants deficient in polyunsaturated fatty acid synthesis. J. Biol. Chem. 267, 1502-1509.

Murray, M.G., Thompson, W.F. (1980) Rapid isolation of high molecular weight plant DNA. Nucl. Acid Res. 8, 4321-4325.

Okuley, J., Lightner, J., Feldman, K., Yadav, N., Lark, E., Browse, J. (1994) *Arabidopsis* FAD2 gene encodes the enzyme that is essential for polyunsaturated lipid. Plant Cell 6, 147-158.

Sambrook, J., Fritsch, E.F., and Maniatis, T., *Molecular Cloning: a Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, 1989.

Shanklin, J., Whittle, E., Fox, B.G. (1994) Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monoxygenase. Biochemistry 33, 12787-12794.

Smith, C.R., Jr. (1985) Unusual seed oils and their fatty acids, in *Fatty Acids*, Pryde E.H., Ed., American Oil Chemists' Society, Champaign, Second edition, pp 29-47.

Töpfer, R., Martini, N., Schell, J. (1995) Modification of plant lipid synthesis. Science 268, 681-686.

van de Loo, F.J., Fox, B.G., Somerville, C. (1993) Unusual fatty acids, in *Lipid Metabolism in Plants*, T.S. Moore Jr., Ed., CRC Press, Boca Raton, pp 91-126.

van de Loo, F.N., Turner, S., Somerville, C.R. (1995) An oleate 12-hydroxylase from castor (*Ricinus communis* L.) is a fatty acyl desaturase homolog. Proc. Natl. Acad. Sci. USA 92, 6743-6747.

von Heijne, G. (1985) Signal sequences. J. Mol. Biol. 184, 99-105.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Somerville, Chris
Broun, Pierre
van de Loo, Frank
Boddupalli, Sekhar S.

(ii) TITLE OF INVENTION: Production of Hydroxylated Fatty Acids in Genetically Modified Plants

(iii) NUMBER OF SEQUENCES: 15

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: PILLSBURY MADISON & SUTRO
(B) STREET: 1100 NEW YORK AVENUE, N.W.
(C) CITY: WASHINGTON
(D) STATE: D.C.
(E) COUNTRY: USA
(F) ZIP: 20005-3918

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: 3.5 inch, 1.44 MB storage
(B) COMPUTER: IBM compatible
(C) OPERATING SYSTEM: DOS 5.0
(D) SOFTWARE: Word Perfect 5.1

(vi) CURRENT APPLICATION DATA;
(A) APPLICATION NUMBER: not yet assigned
(B) FILING DATE: February 6, 1997
(C) CLASSIFICATION:

(2) INFORMATION FOR SEQ ID NO:1

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 543 nucleotides
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TATTGGCACC GGCGGCACCA TTCCAACAAT GGATCCCTAG 40

AAAAAGATGA AGTCTTTGTC CCACCTAAGA AAGCTGCAGT 80

CANATGGTAT GTCAAATACC TCAACAACCC TCTTGGACGC 120

ATTCTGGTGT TAACAGTTCA GTTTATCCTC GGGTGGCCTT 160

TGTATCTAGC CTTTAATGTA TCAGGTAGAC CTTATGATGG 200

TTTCGCTTCA CATTTCTTCC CTCATGCACC TATCTTTAAG 240

GACCGTGAAC GTCTCCAGAT ATACATCTCA GATGCTGGTA 280

TTCTAGCTGT CTGTTATGGT CTTTACCGTT ACGCTGCTTC 320

ACAAGGATTG ACTGCTATGA TCTGCGTCTA CGGAGTACCG 360

CTTTTGATAG TGAACTTTTT CCTTGTCTTG GTCACTTTCT 400

TGCAGCACAC TCATCCTTCA TTACCTCACT ATGATTCAAC 440

CGAGTGGGAA TGGATTAGAG GAGCTTTGGT TACGGTAGAC 480

AGAGACTATG GAATCTTGAA CAAGGTGTTT CACAACATAA 520

CAGACACCCA CGTAGCACAC CAC 543

(2) INFORMATION FOR SEQ ID NO:2

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 544 nucleotides
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TATAGGCACC GGAGGCACCA TTCCAACACA GGATCCCTCG 40

AAAGAGATGA AGTATTTGTC CCAAAGCAGA AATCCGCAAT 80

CAAGTGGTAC GGCGAATACC TCAACAACCC TCCTGGTCGC 120

ATCATGATGT TAACTGTCCA GTTCGTCCTC GGATGGCCCT 160

TGTACTTAGC CTTCAACGTT TCTGGCAGAC CCTACAATGG 200

TTTCGCTTCC CATTTCTTCC CCAATGCTCC TATCTACAAC 240

GACCGTGAAC GCCTCCAGAT TTACATCTCT GATGCTGGTA 280

TTCTAGCCGT CTGTTATGGT CTTTACCGTT ACGCTGTTGC 320

ACAAGGACTA GCCTCAATGA TCTGTCTAAA CGGAGTTCCG 360

CTTCTGATAG TTAACTTTTT CCTCGTCTTG ATCACTTACT 400

TACAACACAC TCACCCTGCG TTGCCTCACT ATGATTCATC 440

AGAGTGGGAT TGGCTTAGAG GAGCTTTAGC TACTGTAGAC 480

AGAGACTATG GAATCTTGAA CAAGGTGTTC CATAACATCA 520

CAGACACCCA CGTCGCACAC CACT 544

(2) INFORMATION FOR SEQ ID NO:3

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1855 nucleotides
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGAAGCTTT ATAAGAAGTT AGTTTTCTCT GGTGACAGAG 40

AAATTNTGTC AATTGGTAGT GACAGTTGAA GCAACAGGAA 80

CAACAAGGAT GGTTGGTGNT GATGCTGATG TGGTGATGTG 120

TTATTCATCA AATACTAAAT ACTACATTAC TTGTTGCTGC 160

CTACTTCTCC TATTTCCTCC GCCACCCATT TTGGACCCAC 200

GANCCTTCCA TTTAAACCCT CTCTCGTGCT ATTCACCAGA 240

AGAGAAGCCA AGAGAGAGAG AGAGAGAATG TTCTGAGGAT 280

CATTGTCTTC TTCATCGTTA TTAACGTAAG TTTTTTTTGA 320

CCACTCATAT CTAAAATCTA GTACATGCAA TAGATTAATG 360

ACTGTTCCTT CTTTTGATAT TTTCAGCTTC TTGAATTCAA 400

GATGGGTGCT GGTGGAAGAA TAATGGTTAC CCCCTCTTCC 440

AAGAAATCAG AAACTGAAGC CCTAAAACGT GGACCATGTG 480

AGAAACCACC ATTCACTGTT AAAGATCTGA AGAAAGCAAT 520

CCCACAGCAT TGTTTCAAGC GCTCTATCCC TCGTTCTTTC 560

TCCTACCTTC TCACAGATAT CACTTTAGTT TCTTGCTTCT 600

ACTACGTTGC CACAAATTAC TTCTCTCTTC TTCCTCAGCC 640

TCTCTCTACT TACCTAGCTT GGCCTCTCTA TTGGGTATGT 680
CAAGGCTGTG TCTTAACCGG TATCTGGGTC ATTGGCCATG 720
AATGTGGTCA CCATGCATTC AGTGACTATC AATGGGTAGA 760
TGACACTGTT GGTTTTATCT TCCATTCCTT CCTTCTCGTC 800
CCTTACTTCT CCTGGAAATA CAGTCATCGT CGTCACCATT 840
CCAACAATGG ATCTCTCGAG AAAGATGAAG TCTTTGTCCC 880
ACCGAAGAAA GCTGCAGTCA AATGGTATGT TAAATACCTC 920
AACAACCCTC TTGGACGCAT TCTGGTGTTA ACAGTTCAGT 960
TTATCCTCGG GTGGCCTTTG TATCTAGCCT TTAATGTATC 1000
AGGTAGACCT TATGATGGTT TCGCTTCACA TTTCTTCCCT 1040
CATGCACCTA TCTTTAAAGA CCGAGAACGC CTCCAGATAT 1080
ACATCTCAGA TGCTGGTATT CTAGCTGTCT GTTATGGTCT 1120
TTACCGTTAC GCTGCTTCAC AAGGATTGAC TGCTATGATC 1160
TGCGTCTATG GAGTACCGCT TTTGATAGTG AACTTTTTCC 1200
TTGTCTTGGT AACTTTCTTG CAGCACACTC ATCCTTCGTT 1240
ACCTCATTAT GATTCAACCG AGTGGGAATG GATTAGAGGA 1280
GCTTTGGTTA CGGTAGACAG AGACTATGGA ATATTGAACA 1320
AGGTGTTCCA TAACATAACA GACACACATG TGGCTCATCA 1360
TCTCTTTGCA ACTATACCGC ATTATAACGC AATGGAAGCT 1400
ACAGAGGCGA TAAAGCCAAT ACTTGGTGAT TACTACCACT 1440
TCGATGGAAC ACCGTGGTAT GTGGCCATGT ATAGGGAAGC 1480
AAAGGAGTGT CTCTATGTAG AACCGGATAC GGAACGTGGG 1520
AAGAAAGGTG TCTACTATTA CAACAATAAG TTATGAGGCT 1560
GATAGGGCGA GAGAAGTGCA ATTATCAATC TTCATTTCCA 1600
TGTTTTAGGT GTCTTGTTTA AGAAGCTATG CTTTGTTTCA 1640
ATAATCTCAG AGTCCATNTA GTTGTGTTCT GGTGCATTTT 1680

GCCTAGTTAT GTGGTGTCGG AAGTTAGTGT TCAAACTGCT 1720

TCCTGCTGTG CTGCCCAGTG AAGAACAAGT TTACGTGTTT 1760

AAAATACTCG GAACGAATTG ACCACAANAT ATCCAAAACC 1800

GGCTATCCGA ATTCCATATC CGAAAACCGG ATATCCAAAT 1840

TTCCAGAGTA CTTAG 1855

(2) INFORMATION FOR SEQ ID NO:4

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 384 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Gly Ala Gly Gly Arg Ile Met Val Thr
5 10

Pro Ser Ser Lys Lys Ser Glu Thr Glu Ala
15 20

Leu Lys Arg Gly Pro Cys Glu Lys Pro Pro
25 30

Phe Thr Val Lys Asp Leu Lys Lys Ala Ile
35 40

Pro Gln His Cys Phe Lys Arg Ser Ile Pro
45 50

Arg Ser Phe Ser Tyr Leu Leu Thr Asp Ile
55 60

Thr Leu Val Ser Cys Phe Tyr Tyr Val Ala
65 70

Thr Asn Tyr Phe Ser Leu Leu Pro Gln Pro
75 80

Leu Ser Thr Tyr Leu Ala Trp Pro Leu Tyr
85 90

Trp Val Cys Gln Gly Cys Val Leu Thr Gly
95 100

Ile Trp Val Ile Gly His Glu Cys Gly His
105 110

His Ala Phe Ser Asp Tyr Gln Trp Val Asp
115 120

Asp Thr Val Gly Phe Ile Phe His Ser Phe
125 130

Leu Leu Val Pro Tyr Phe Ser Trp Lys Tyr
135 140

Ser His Arg Arg His His Ser Asn Asn Gly
145 150

Ser Leu Glu Lys Asp Glu Val Phe Val Pro
155 160

Pro Lys Lys Ala Ala Val Lys Trp Tyr Val
165 170

Lys Tyr Leu Asn Asn Pro Leu Gly Arg Ile
175 180

Leu Val Leu Thr Val Gln Phe Ile Leu Gly
185 190

Trp Pro Leu Tyr Leu Ala Phe Asn Val Ser
195 200

Gly Arg Pro Tyr Asp Gly Phe Ala Ser His
205 210

Phe Phe Pro His Ala Pro Ile Phe Lys Asp
215 220

Arg Glu Arg Leu Gln Ile Tyr Ile Ser Asp
225 230

Ala Gly Ile Leu Ala Val Cys Tyr Gly Leu
235 240

Tyr Arg Tyr Ala Ala Ser Gln Gly Leu Thr
245 250

Ala Met Ile Cys Val Tyr Gly Val Pro Leu
255 260

Leu Ile Val Asn Phe Phe Leu Val Leu Val
265 270

Thr Phe Leu Gln His Thr His Pro Ser Leu
275 280

Pro His Tyr Asp Ser Thr Glu Trp Glu Trp
285 290

Ile Arg Gly Ala Leu Val Thr Val Asp Arg
295 300

Asp Tyr Gly Ile Leu Asn Lys Val Phe His
305 310

Asn Ile Thr Asp Thr His Val Ala His His
315 320

Leu Phe Ala Thr Ile Pro His Tyr Asn Ala
325 330

Met Glu Ala Thr Glu Ala Ile Lys Pro Ile
335 340

Leu Gly Asp Tyr Tyr His Phe Asp Gly Thr
345 350

Pro Trp Tyr Val Ala Met Tyr Arg Glu Ala
355 360

Lys Glu Cys Leu Tyr Val Glu Pro Asp Thr
365 370

Glu Arg Gly Lys Lys Gly Val Tyr Tyr Tyr
375 380

Asn Asn Lys Leu

(2) INFORMATION FOR SEQ ID NO:5

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 387 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Gly Gly Gly Gly Arg Met Ser Thr Val
5 10

Ile Thr Ser Asn Asn Ser Glu Lys Lys Gly
15 20

Gly Ser Ser His Leu Lys Arg Ala Pro His
25 30

Thr Lys Pro Pro Phe Thr Leu Gly Asp Leu
35 40

Lys Arg Ala Ile Pro Pro His Cys Phe Glu
45 50

Arg Ser Phe Val Arg Ser Phe Ser Tyr Val
55 60

Ala Tyr Asp Val Cys Leu Ser Phe Leu Phe
65 70

Tyr Ser Ile Ala Thr Asn Phe Phe Pro Tyr
75 80

Ile Ser Ser Pro Leu Ser Tyr Val Ala Trp
85 90

Leu Val Tyr Trp Leu Phe Gln Gly Cys Ile
95 100

Leu Thr Gly Leu Trp Val Ile Gly His Glu
105 110

Cys Gly His His Ala Phe Ser Glu Tyr Gln
115 120

Leu Ala Asp Asp Ile Val Gly Leu Ile Val
125 130

His Ser Ala Leu Leu Val Pro Tyr Phe Ser
135 140

Trp Lys Tyr Ser His Arg Arg His His Ser
145 150

Asn Ile Gly Ser Leu Glu Arg Asp Glu Val
155 160

Phe Val Pro Lys Ser Lys Ser Lys Ile Ser
165 170

Trp Tyr Ser Lys Tyr Ser Asn Asn Pro Pro
175 180

Gly Arg Val Leu Thr Leu Ala Ala Thr Leu
185 190

Leu Leu Gly Trp Pro Leu Tyr Leu Ala Phe
195 200

Asn Val Ser Gly Arg Pro Tyr Asp Arg Phe
205 210

Ala Cys His Tyr Asp Pro Tyr Gly Pro Ile
215 220

Phe Ser Glu Arg Glu Arg Leu Gln Ile Tyr
225 230

Ile Ala Asp Leu Gly Ile Phe Ala Thr Thr
235 240

Phe Val Leu Tyr Gln Ala Thr Met Ala Lys
245 250

Gly Leu Ala Trp Val Met Arg Ile Tyr Gly
255 260

Val Pro Leu Leu Ile Val Asn Cys Phe Leu
265 270

Val Met Ile Thr Tyr Leu Gln His Thr His
275 280

Pro Ala Ile Pro Arg Tyr Gly Ser Ser Glu
285 290

Trp Asp Trp Leu Arg Gly Ala Met Val Thr
295 300

Val Asp Arg Asp Tyr Gly Val Leu Asn Lys
305 310

Val Phe His Asn Ile Ala Asp Thr His Val
315 320

Ala His His Leu Phe Ala Thr Val Pro His
325 330

Tyr His Ala Met Glu Ala Thr Lys Ala Ile
335 340

Lys Pro Ile Met Gly Glu Tyr Tyr Arg Tyr
345 350

Asp Gly Thr Pro Phe Tyr Lys Ala Leu Trp
355 360

Arg Glu Ala Lys Glu Cys Leu Phe Val Glu
365 370

Pro Asp Glu Gly Ala Pro Thr Gln Gly Val
375 380

Phe Trp Tyr Arg Asn Lys Tyr
385

(2) INFORMATION FOR SEQ ID NO:6

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 383 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Gly Ala Gly Gly Arg Met Pro Val Pro
5 10

Thr Ser Ser Lys Lys Ser Glu Thr Asp Thr
15 20

Thr Lys Arg Val Pro Cys Glu Lys Pro Pro
25 30

Phe Ser Val Gly Asp Leu Lys Lys Ala Ile
35 40

Pro Pro His Cys Phe Lys Arg Ser Ile Pro
45 50

Arg Ser Phe Ser Tyr Leu Ile Ser Asp Ile
55 60

Ile Ile Ala Ser Cys Phe Tyr Tyr Val Ala
65 70

Thr Asn Tyr Phe Ser Leu Leu Pro Gln Pro
75 80

Leu Ser Tyr Leu Ala Trp Pro Leu Tyr Trp
85 90

Ala Cys Gln Gly Cys Val Leu Thr Gly Ile
95 100

Trp Val Ile Ala His Glu Cys Gly His His
105 110

Ala	Phe	Ser	Asp	Tyr 115	Gln	Trp	Leu	Asp	Asp 120
Thr	Val	Gly	Leu	Ile 125	Phe	His	Ser	Phe	Leu 130
Leu	Val	Pro	Tyr	Phe 135	Ser	Trp	Lys	Tyr	Ser 140
His	Arg	Arg	His	His 145	Ser	Asn	Thr	Gly	Ser 150
Leu	Glu	Arg	Asp	Glu 155	Val	Phe	Val	Pro	Lys 160
Gln	Lys	Ser	Ala	Ile 165	Lys	Trp	Tyr	Gly	Lys 170
Tyr	Leu	Asn	Asn	Pro 175	Leu	Gly	Arg	Ile	Met 180
Met	Leu	Thr	Val	Gln 185	Phe	Val	Leu	Gly	Trp 190
Pro	Leu	Tyr	Leu	Ala 195	Phe	Asn	Val	Ser	Gly 200
Arg	Pro	Tyr	Asp	Gly 205	Phe	Ala	Cys	His	Phe 210
Phe	Pro	Asn	Ala	Pro 215	Ile	Tyr	Asn	Asp	Arg 220
Glu	Arg	Leu	Gln	Ile 225	Tyr	Leu	Ser	Asp	Ala 230
Gly	Ile	Leu	Ala	Val 235	Cys	Phe	Gly	Leu	Tyr 240
Arg	Tyr	Ala	Ala	Ala 245	Gln	Gly	Met	Ala	Ser 250
Met	Ile	Cys	Leu	Tyr 255	Gly	Val	Pro	Leu	Leu 260
Ile	Val	Asn	Ala	Phe 265	Leu	Val	Leu	Ile	Thr 270
Tyr	Leu	Gln	His	Thr 275	His	Pro	Ser	Leu	Pro 280

His Tyr Asp Ser Ser Glu Trp Asp Trp Leu
285 290

Arg Gly Ala Leu Ala Thr Val Asp Arg Asp
295 300

Tyr Gly Ile Leu Asn Lys Val Phe His Asn
305 310

Ile Thr Asp Thr His Val Ala His His Leu
315 320

Phe Ser Thr Met Pro His Tyr Asn Ala Met
325 330

Glu Ala Thr Lys Ala Ile Lys Pro Ile Leu
335 340

Gly Asp Tyr Tyr Gln Phe Asp Gly Thr Pro
345 350

Trp Tyr Val Ala Met Tyr Arg Glu Ala Lys
355 360

Glu Cys Ile Tyr Val Glu Pro Asp Arg Glu
365 370

Gly Asp Lys Lys Gly Val Tyr Trp Tyr Asn
375 380

Asn Lys Leu

(2) INFORMATION FOR SEQ ID NO:7

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 384 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Gly Ala Gly Gly Arg Met Gln Val Ser
5 10

Pro Pro Ser Lys Lys Ser Glu Thr Asp Asn
15 20

Ile Lys Arg Val Pro Cys Glu Thr Pro Pro
25 30

Phe Thr Val Gly Glu Leu Lys Lys Ala Ile
35 40

Pro Pro His Cys Phe Lys Arg Ser Ile Pro
45 50

Arg Ser Phe Ser His Leu Ile Trp Asp Ile
55 60

Ile Ile Ala Ser Cys Phe Tyr Tyr Val Ala
65 70

Thr Thr Tyr Phe Pro Leu Leu Pro Asn Pro
75 80

Leu Ser Tyr Phe Ala Trp Pro Leu Tyr Trp
85 90

Ala Cys Gln Gly Cys Val Leu Thr Gly Val
95 100

Trp Val Ile Ala His Glu Cys Gly His Ala
105 110

Ala Phe Ser Asp Tyr Gln Trp Leu Asp Asp
115 120

Thr Val Gly Leu Ile Phe His Ser Phe Leu
125 130

Leu Val Pro Tyr Phe Ser Trp Lys Tyr Ser
135 140

His Arg Arg His His Ser Asn Thr Gly Ser
145 150

Leu Glu Arg Asp Glu Val Phe Val Pro Arg
155 160

Arg Ser Gln Thr Ser Ser Gly Thr Ala Ser
165 170

Thr Ser Thr Thr Phe Gly Arg Thr Val Met
175 180

Leu Thr Val Gln Phe Thr Leu Gly Trp Pro
185 190

Leu Tyr Leu Ala Phe Asn Val Ser Gly Arg
195 200

Pro Tyr Asp Gly Gly Phe Ala Cys His Phe
205 210

His Pro Asn Ala Pro Ile Tyr Asn Asp Arg
215 220

Glu Arg Leu Gln Ile Tyr Ile Ser Asp Ala
225 230

Gly Ile Leu Ala Val Cys Tyr Gly Leu Leu
235 240

Pro Tyr Ala Ala Val Gln Gly Val Ala Ser
245 250

Met Val Cys Phe Leu Arg Val Pro Leu Leu
255 260

Ile Val Asn Gly Phe Leu Val Leu Ile Thr
265 270

Tyr Leu Gln His Thr His Pro Ser Leu Pro
275 280

His Tyr Asp Ser Ser Glu Trp Asp Trp Leu
285 290

Arg Gly Ala Leu Ala Thr Val Asp Arg Asp
295 300

Tyr Gly Ile Leu Asn Gln Gly Phe His Asn
305 310

Ile Thr Asp Thr His Glu Ala His His Leu
315 320

Phe Ser Thr Met Pro His Tyr His Ala Met
325 330

Glu Ala Thr Lys Ala Ile Lys Pro Ile Leu
335 340

Gly Glu Tyr Tyr Gln Phe Asp Gly Thr Pro
345 350

Val Val Lys Ala Met Trp Arg Glu Ala Lys
355 360

Glu Cys Ile Tyr Val Glu Pro Asp Arg Gln
365 370

```
Gly Glu Lys Lys Gly Val Phe Trp Tyr Asn
                375                 380

Asn Lys Leu Xaa
```

(2) INFORMATION FOR SEQ ID NO:8

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 309 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Ser Leu Leu Thr Ser Phe Ser Tyr Val Val
                5                   10

Tyr Asp Leu Ser Phe Ala Phe Ile Phe Tyr
                15                  20

Ile Ala Thr Thr Tyr Phe His Leu Leu Pro
                25                  30

Gln Pro Phe Ser Leu Ile Ala Trp Pro Ile
                35                  40

Tyr Trp Val Leu Gln Gly Cys Leu Leu Thr
                45                  50

Arg Val Cys Gly His His Ala Phe Ser Lys
                55                  60

Tyr Gln Trp Val Asp Asp Val Val Gly Leu
                65                  70

Thr Leu His Ser Thr Leu Leu Val Pro Tyr
                75                  80

Phe Ser Trp Lys Ile Ser His Arg Arg His
                85                  90

His Ser Asn Thr Gly Ser Leu Asp Arg Asp
                95                  100

Glu Arg Val Lys Val Ala Trp Phe Ser Lys
                105                 110

Tyr Leu Asn Asn Pro Leu Gly Arg Ala Val
                115                 120
```

```
Ser Leu Leu Val Thr Leu Thr Ile Gly Trp
                125                 130

Pro Met Tyr Leu Ala Phe Asn Val Ser Gly
                135                 140

Arg Pro Tyr Asp Ser Phe Ala Ser His Tyr
                145                 150

His Pro Tyr Arg Val Arg Leu Leu Ile Tyr
                155                 160

Val Ser Asp Val Ala Leu Phe Ser Val Thr
                165                 170

Tyr Ser Leu Tyr Arg Val Ala Thr Leu Lys
                175                 180

Gly Leu Val Trp Leu Leu Cys Val Tyr Gly
                185                 190

Val Pro Leu Leu Ile Val Asn Gly Phe Leu
                195                 200

Val Thr Ile Thr Tyr Leu Arg Val His Tyr
                205                 210

Asp Ser Ser Glu Trp Asp Trp Leu Lys Gly
                215                 220

Ala Leu Ala Thr Met Asp Arg Asp Tyr Gly
                225                 230

Ile Leu Asn Lys Val Phe His His Ile Thr
                235                 240

Asp Thr His Val Ala His His Leu Phe Ser
                245                 250

Thr Met Pro His Tyr His Leu Arg Val Lys
                255                 260

Pro Ile Leu Gly Glu Tyr Tyr Gln Phe Asp
                265                 270

Asp Thr Pro Phe Tyr Lys Ala Leu Trp Arg
                275                 280

Glu Ala Arg Glu Cys Leu Tyr Val Glu Pro
                285                 290
```

Asp Glu Gly Thr Ser Glu Lys Gly Val Tyr
295 300

Trp Tyr Arg Asn Lys Tyr Leu Arg Val
305

(2) INFORMATION FOR SEQ ID NO:9

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 302 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Phe Ser Tyr Val Val Tyr Asp Leu Thr Ile
5 10

Ala Phe Cys Leu Tyr Tyr Val Ala Thr His
15 20

Tyr Phe His Leu Leu Pro Gly Pro Leu Ser
25 30

Phe Arg Gly Met Ala Ile Tyr Trp Ala Val
35 40

Gln Gly Cys Ile Leu Thr Gly Val Trp Val
45 50

Val Ala Phe Ser Asp Tyr Gln Leu Leu Asp
55 60

Asp Ile Val Gly Leu Ile Leu His Ser Ala
65 70

Leu Leu Val Pro Tyr Phe Ser Trp Lys Tyr
75 80

Ser His Arg Arg His His Ser Asn Thr Gly
85 90

Ser Leu Glu Arg Asp Glu Val Phe Val Pro
95 100

Lys Val Ser Lys Tyr Leu Asn Asn Pro Pro
105 110

Gly Arg Val Leu Thr Leu Ala Val Thr Leu
115 120

Thr Leu Gly Trp Pro Leu Tyr Leu Ala Leu
125 130

Asn Val Ser Gly Arg Pro Tyr Asp Arg Phe
135 140

Ala Cys His Tyr Asp Pro Tyr Gly Pro Ile
145 150

Tyr Ser Val Ile Ser Asp Ala Gly Val Leu
155 160

Ala Val Val Tyr Gly Leu Phe Arg Leu Ala
165 170

Met Ala Lys Gly Leu Ala Trp Val Val Cys
175 180

Val Tyr Gly Val Pro Leu Leu Val Val Asn
185 190

Gly Phe Leu Val Leu Ile Thr Phe Leu Gln
195 200

His Thr His Val Ser Glu Trp Asp Trp Leu
205 210

Arg Gly Ala Leu Ala Thr Val Asp Arg Asp
215 220

Tyr Gly Ile Leu Asn Lys Val Phe His Asn
225 230

Ile Thr Asp Thr His Val Ala His His Leu
235 240

Phe Ser Thr Met Pro His Tyr His Ala Met
245 250

Glu Ala Thr Val Glu Tyr Tyr Arg Phe Asp
255 260

Glu Thr Pro Phe Val Lys Ala Met Trp Arg
265 270

Glu Ala Arg Glu Cys Ile Tyr Val Glu Pro
275 280

Asp Gln Ser Thr Glu Ser Lys Gly Val Phe
285 290

Trp Tyr Asn Asn Lys Leu Ala Met Glu Ala
295 300

Thr Val

(2) INFORMATION FOR SEQ ID NO:10

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 372 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Gly Ala Gly Gly Arg Met Thr Glu Lys
5 10

Glu Arg Glu Lys Gln Glu Gln Leu Ala Arg
15 20

Ala Thr Gly Gly Ala Ala Met Gln Arg Ser
25 30

Pro Val Glu Lys Pro Pro Phe Thr Leu Gly
35 40

Gln Ile Lys Lys Ala Ile Pro Pro His Cys
45 50

Phe Glu Arg Ser Val Leu Lys Ser Phe Ser
55 60

Tyr Val Val His Asp Leu Val Ile Ala Ala
65 70

Ala Leu Leu Tyr Phe Ala Leu Ala Ile Ile
75 80

Pro Ala Leu Pro Ser Pro Leu Arg Tyr Ala
85 90

Ala Trp Pro Leu Tyr Trp Ile Ala Gln Gly
95 100

Ala Phe Ser Asp Tyr Ser Leu Leu Asp Asp
105 110

Val Val Gly Leu Val Leu His Ser Ser Leu
115 120

```
Met Val Pro Tyr Phe Ser Trp Lys Tyr Ser
                125                 130

His Arg Arg His His Ser Asn Thr Gly Ser
                135                 140

Leu Glu Arg Asp Glu Val Phe Val Pro Lys
                145                 150

Lys Lys Glu Ala Leu Pro Trp Tyr Thr Pro
                155                 160

Tyr Val Tyr Asn Asn Pro Val Gly Arg Val
                165                 170

Val His Ile Val Val Gln Leu Thr Leu Gly
                175                 180

Trp Pro Leu Tyr Leu Ala Thr Asn Ala Ser
                185                 190

Gly Arg Pro Tyr Pro Arg Phe Ala Cys His
                195                 200

Phe Asp Pro Tyr Gly Pro Ile Tyr Asn Asp
                205                 210

Arg Glu Arg Ala Gln Ile Phe Val Ser Asp
                215                 220

Ala Gly Val Val Ala Val Ala Phe Gly Leu
                225                 230

Tyr Lys Leu Ala Ala Ala Phe Gly Val Trp
                235                 240

Trp Val Val Arg Val Tyr Ala Val Pro Leu
                245                 250

Leu Ile Val Asn Ala Trp Leu Val Leu Ile
                255                 260

Thr Tyr Leu Gln His Thr His Pro Ser Leu
                265                 270

Pro His Tyr Asp Ser Ser Glu Trp Asp Trp
                275                 280

Leu Arg Gly Ala Leu Ala Thr Met Asp Arg
                285                 290
```

Asp Tyr Gly Ile Leu Asn Arg Val Phe His
295 300

Asn Ile Thr Asp Thr His Val Ala His His
305 310

Leu Phe Ser Thr Met Pro His Tyr His Ala
315 320

Met Glu Ala Thr Lys Ala Ile Arg Pro Ile
325 330

Leu Gly Asp Tyr Tyr His Phe Asp Pro Thr
335 340

Pro Val Ala Lys Ala Thr Trp Arg Glu Ala
345 350

Gly Glu Cys Ile Tyr Val Glu Pro Glu Asp
355 360

Arg Lys Gly Val Phe Trp Tyr Asn Lys Lys
365 370

Phe Xaa

(2) INFORMATION FOR SEQ ID NO:11

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 224 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Trp Val Met Ala His Asp Cys Gly His His
5 10

Ala Phe Ser Asp Tyr Gln Leu Leu Asp Asp
15 20

Val Val Gly Leu Ile Leu His Ser Cys Leu
25 30

Leu Val Pro Tyr Phe Ser Trp Lys His Ser
35 40

His Arg Arg His His Ser Asn Thr Gly Ser
45 50

Leu Glu Arg Asp Glu Val Phe Val Pro Lys
55 60

Lys Lys Ser Ser Ile Arg Trp Tyr Ser Lys
65 70

Tyr Leu Asn Asn Pro Pro Gly Arg Ile Met
75 80

Thr Ile Ala Val Thr Leu Ser Leu Gly Trp
85 90

Pro Leu Tyr Leu Ala Phe Asn Val Ser Gly
95 100

Arg Pro Tyr Asp Arg Phe Ala Cys His Tyr
105 110

Asp Pro Tyr Gly Pro Ile Tyr Asn Asp Arg
115 120

Glu Arg Ile Glu Ile Phe Ile Ser Asp Ala
125 130

Gly Val Leu Ala Val Thr Phe Gly Leu Tyr
135 140

Gln Leu Ala Ile Ala Lys Gly Leu Ala Trp
145 150

Val Val Cys Val Tyr Gly Val Pro Leu Leu
155 160

Val Val Asn Ser Phe Leu Val Leu Ile Thr
165 170

Phe Leu Gln His Thr His Pro Ala Leu Pro
175 180

His Tyr Asp Ser Ser Glu Trp Asp Trp Leu
185 190

Arg Gly Ala Leu Ala Thr Val Asp Arg Asp
195 200

Tyr Gly Ile Leu Asn Lys Val Phe His Asn
205 210

Ile Thr Asp Thr Gln Val Ala His His Leu
215 220

Phe Thr Met Pro

(2) INFORMATION FOR SEQ ID NO:12

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 nucleotides
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCTCTTTTGT GCGCTCATTC 20

(2) INFORMATION FOR SEQ ID NO:13

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 nucleotides
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGGTACCAGA AAACGCCTTG 20

(2) INFORMATION FOR SEQ ID NO:14

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 nucleotides
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TAYWSNCAYM GNMGNCAYCA 20

(2) INFORMATION FOR SEQ ID NO:15

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 nucleotides
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

RTGRTGNGCN ACRTGNGTRT C 21

WHAT IS CLAIMED IS:

1. A method of altering an amount of an unsaturated fatty acid in a seed of a plant comprising: decreasing a fatty acid desaturase activity in the seed by genetic manipulation of at least one of fatty acid desaturase or fatty acid hydroxylase.

2. The method of Claim 1, wherein an endogenous gene for said fatty acid hydroxylase is mutated and thereby decreases fatty acid hydroxylase activity in the seed.

3. The method of Claim 1, wherein said plant is transformed with a nucleic acid containing a sequence which encodes a fatty acid hydroxylase or derivative thereof.

4. The method of Claim 3, wherein said derivative is a dominant negative mutant which thereby alters the amount of the unsaturated fatty acid in the seed.

5. The method of Claim 3, wherein said derivative is a mutant fatty acid hydroxylase in which one or more essential histidine residues have been mutated which thereby alters the amount of the unsaturated fatty acid in the seed.

6. The method of Claim 1, wherein an endogenous gene for said fatty acid desaturase is mutated and thereby decreases fatty acid desaturase activity in the seed.

7. The method of Claim 1, wherein said plant is transformed with a nucleic acid containing a sequence which encodes a fatty acid desaturase or derivative thereof.

8. The method of Claim 7, wherein said derivative is a dominant negative mutant which thereby alters the amount of the unsaturated fatty acid in the seed.

9. The method of Claim 7, wherein said derivative is a mutant fatty acid desaturase in which one or more essential histidine residues have been mutated which thereby alters the amount of the unsaturated fatty acid in the seed.

10. The method of Claim 1, wherein said plant is selected from the group consisting of rapeseed, *Crambe, Brassica juncea,* canola, flax, sunflower, safflower, cotton, cuphea, soybean, peanut, coconut, oil palm and corn.

11. A method of altering an amount of a unsaturated fatty acid comprising:

(a) transforming a plant cell with a nucleic acid containing a sequence which encodes a fatty acid hydroxylase or a dominant negative mutant of fatty acid hydroxylase or a dominant negative mutant of fatty acid desaturase,

(b) growing a seed-bearing plant from the transformed plant cell of step (a), and

(c) identifying a seed from the plant of step (b) with the altered amount of the unsaturated fatty acid in the seed.

12. The method of Claim 11, wherein said nucleic acid contains a sequence which encodes the dominant negative mutant of fatty acid hydroxylase in which one or more essential histidine residues have been mutated.

13. The method of Claim 11, wherein said nucleic acid contains a sequence which encodes the dominant negative mutant of fatty acid hydroxylase which thereby alters the amount of the unsaturated fatty acid in the seed.

14. The method of Claim 11, wherein said nucleic acid contains a sequence which encodes the dominant negative mutant of fatty acid desaturase in which one or more essential histidine residues have been mutated.

15. The method of Claim 11, wherein said nucleic acid contains a sequence which encodes the dominant negative mutant of fatty acid desaturase which thereby alters the amount of the unsaturated fatty acid in the seed.

16. The method of Claim 11, wherein said plant is selected from the group consisting of rapeseed, *Crambe*, *Brassica juncea*, canola, flax, sunflower, safflower, cotton, cuphea, soybean, peanut, coconut, oil palm and corn.

17. A recombinant nucleic acid suitable for use in Claim 1, wherein said nucleic acid contains a sequence encoding a fatty acid hydroxylase with an

amino acid identity of 60% or greater to SEQ ID NO:4.

18. The recombinant nucleic acid of Claim 17, wherein the amino acid identity is 90% or greater to SEQ ID NO:4.

19. The recombinant nucleic acid of Claim 17, wherein the amino acid identity is 100% of SEQ ID NO:4.

20. The recombinant nucleic acid of Claim 17, wherein said nucleic acid contains a sequence having a nucleotide identity of 90% or greater to SEQ ID NO:1, 2 or 3.

21. The recombinant nucleic acid of Claim 17, wherein said nucleic acid contains SEQ ID NO:1, 2 or 3.

22. The recombinant nucleic acid of Claim 17, wherein said sequence is obtainable from a plant species producing a hydroxylated fatty acid.

23. A recombinant nucleic acid suitable for use in Claim 1, wherein said nucleic acid contains a sequence encoding at least one of fatty acid desaturase or fatty acid hydroxylase.

24. The recombinant nucleic acid of Claim 23, wherein said sequence is obtainable from *Ricinus communis* (L.) (castor).

25. The recombinant nucleic acid of Claim 23, wherein said sequence is obtainable from *Lesquerella fendleri*.

26. The recombinant nucleic acid of Claim 23, wherein said nucleic acid contains a sequence encoding at least one of fatty acid desaturase or fatty acid hydroxylase in which one or more essential histidine residues have been mutated.

27. The method of Claim 1 further comprising: processing the seed containing the altered amount of the unsaturated fatty acid to obtain oil and/or seed meal.

28. Oil obtained by the method of Claim 27.

29. Seed meal obtained by the method of Claim 27.

30. Plant obtained by the method of Claim 1.

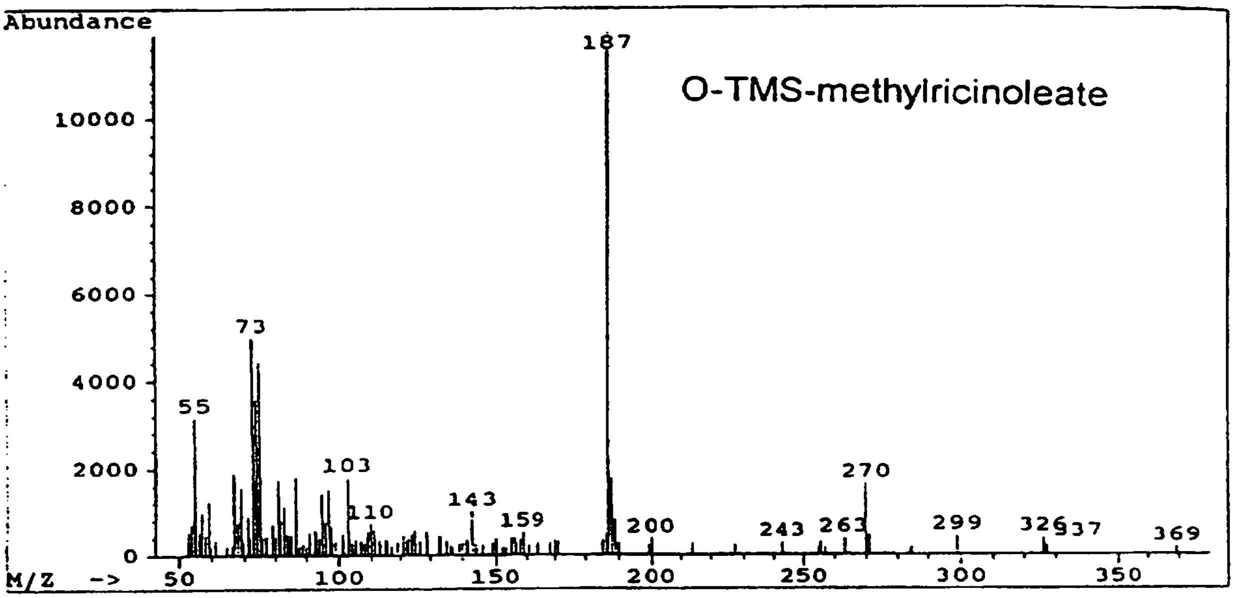
31. The method of Claim 11 further comprising: processing the seed containing the altered amount of the unsaturated fatty acid to obtain oil and/or seed meal.

32. Oil obtained by the method of Claim 31.

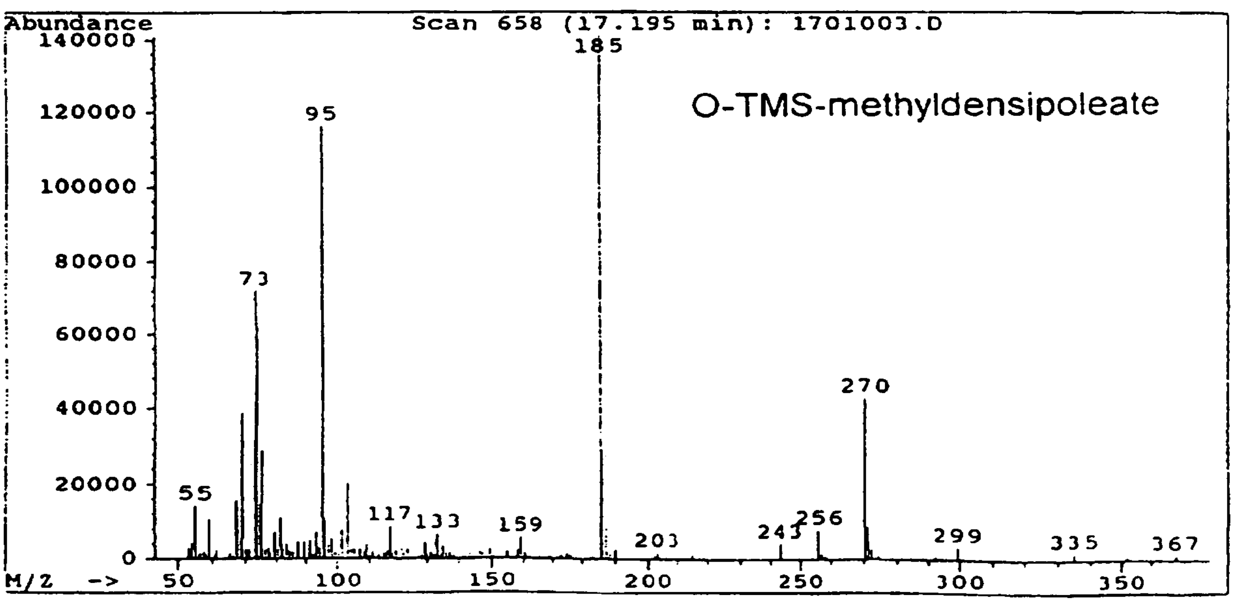
33. Seed meal obtained by the method of Claim 31.

34. Plant obtained by the method of Claim 11.

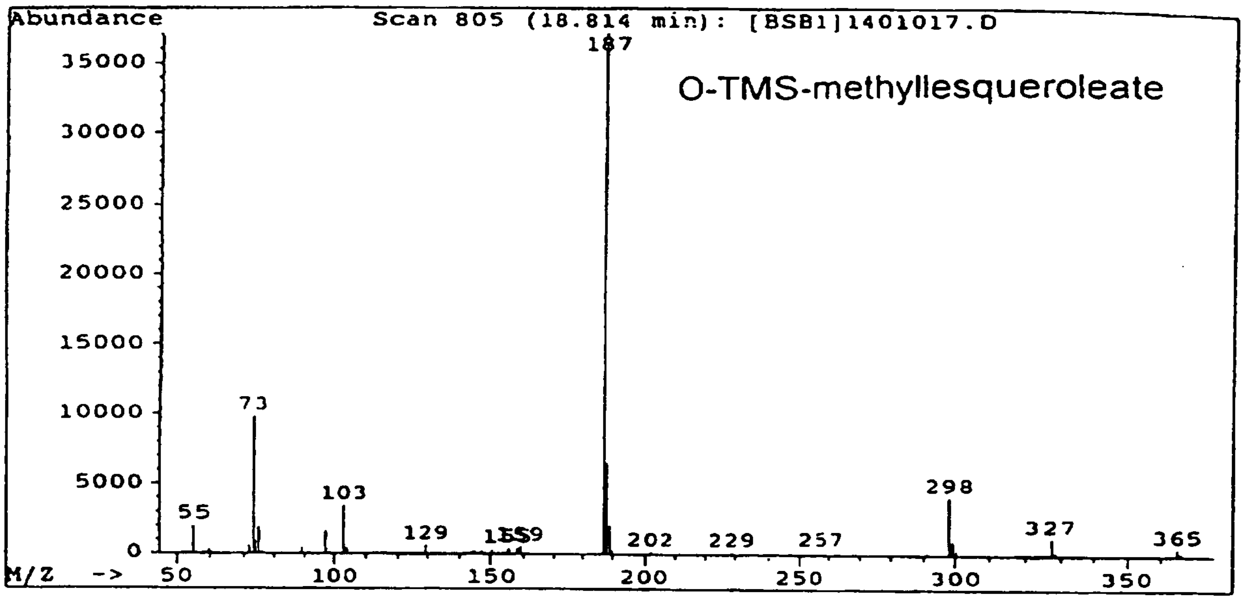
Figure 1A


Abundance
O-TMS-methylricinoleate
10000
8000
6000
4000
2000
0
M/Z ->
50
100
150
200
250
300
350
187
73
55
103
110
143
159
200
243
263
270
299
326
337
369

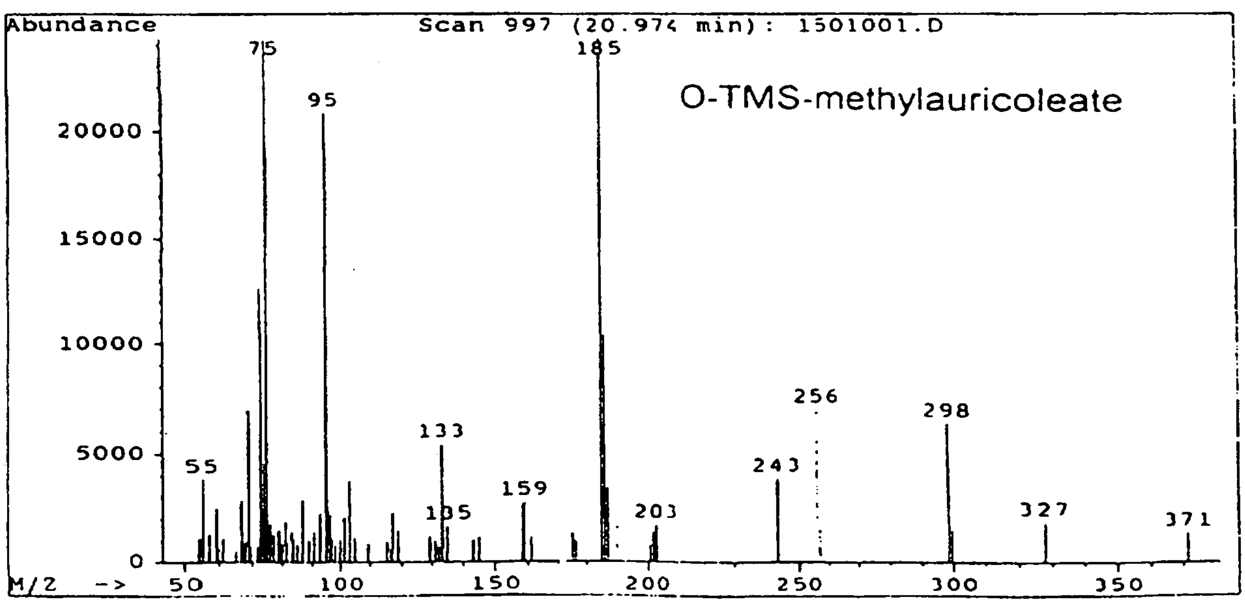
1B


Abundance
Scan 658 (17.195 min): 1701003.D
O-TMS-methyldensipoleate
140000
120000
100000
80000
60000
40000
20000
0
M/Z ->
50
100
150
200
250
300
350
185
95
73
55
117
133
159
203
243
256
270
299
335
367

1C

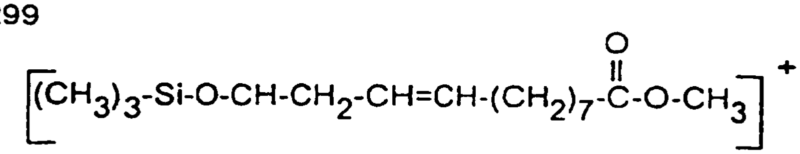

Abundance
Scan 805 (18.814 min): [BSB1]1401017.D
O-TMS-methyllesqueroleate
187
73
103
55
129
159
202
229
257
298
327
365
M/Z ->

1D

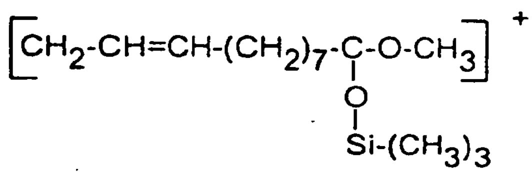

Abundance
Scan 997 (20.974 min): 1501001.D
O-TMS-methylauricoleate
75
185
95
133
135
159
203
243
256
298
327
371
55
M/Z ->

Ion #1: Mass 187 $[CH_3\text{-}(CH_2)_5\text{-}CH\text{-}O\text{-}Si\text{-}(CH_3)_3]^+$

Ion #2: Mass 299



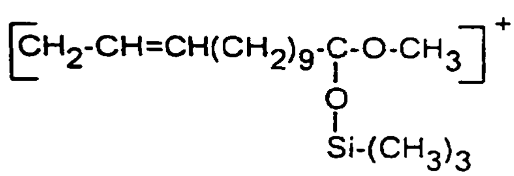
Ion #3: Mass 270 (characteristic rearrangement ion)



Ion #4: Mass 185 (desaturated analog of Ion #1)

$[CH_3\text{-}(CH_2)_2\text{-}CH{=}CH\text{-}CH_2\text{-}CH\text{-}O\text{-}Si\text{-}(CH_3)_3]^+$

Ion #5: Mass 298 (elongated analog of Ion #3)



Ion #6: Mass 327 (elongated analog of ion

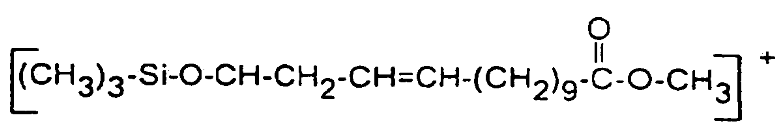


Figure 2

Figure 3

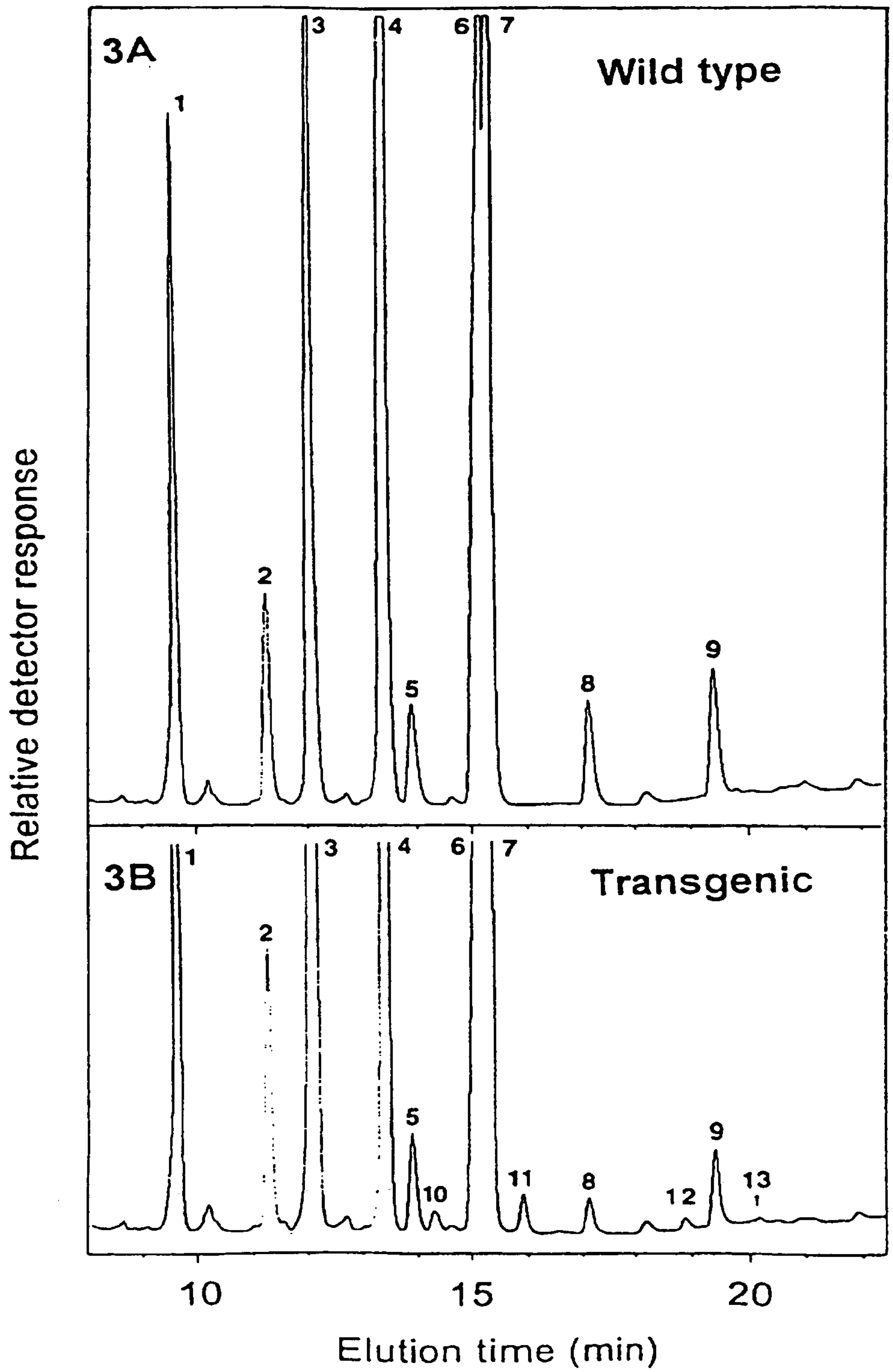
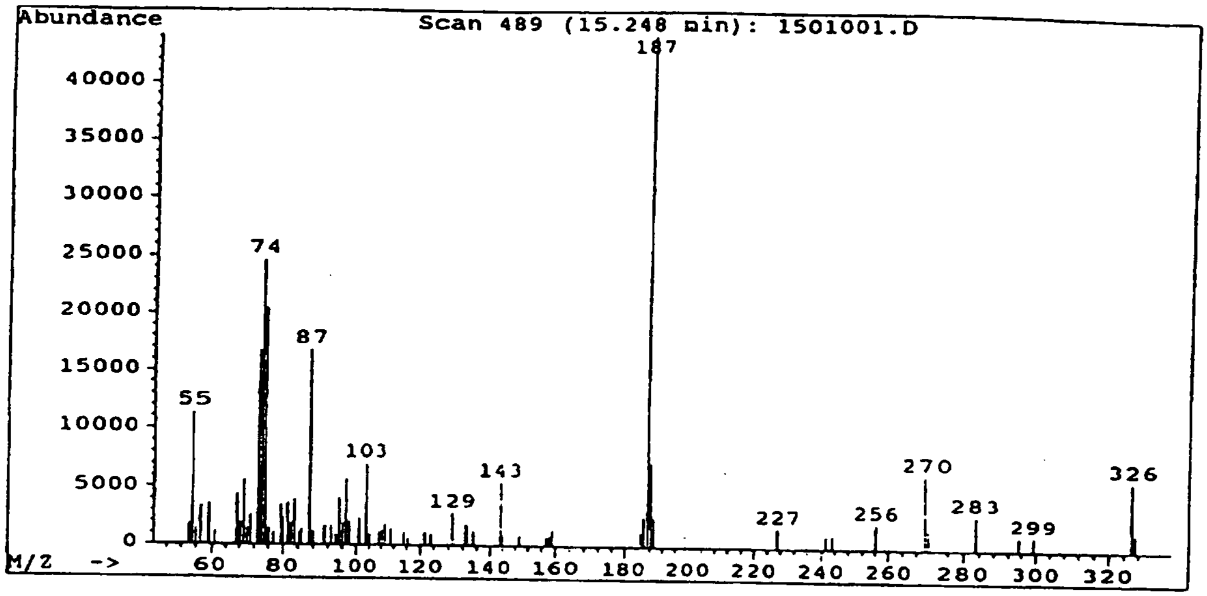
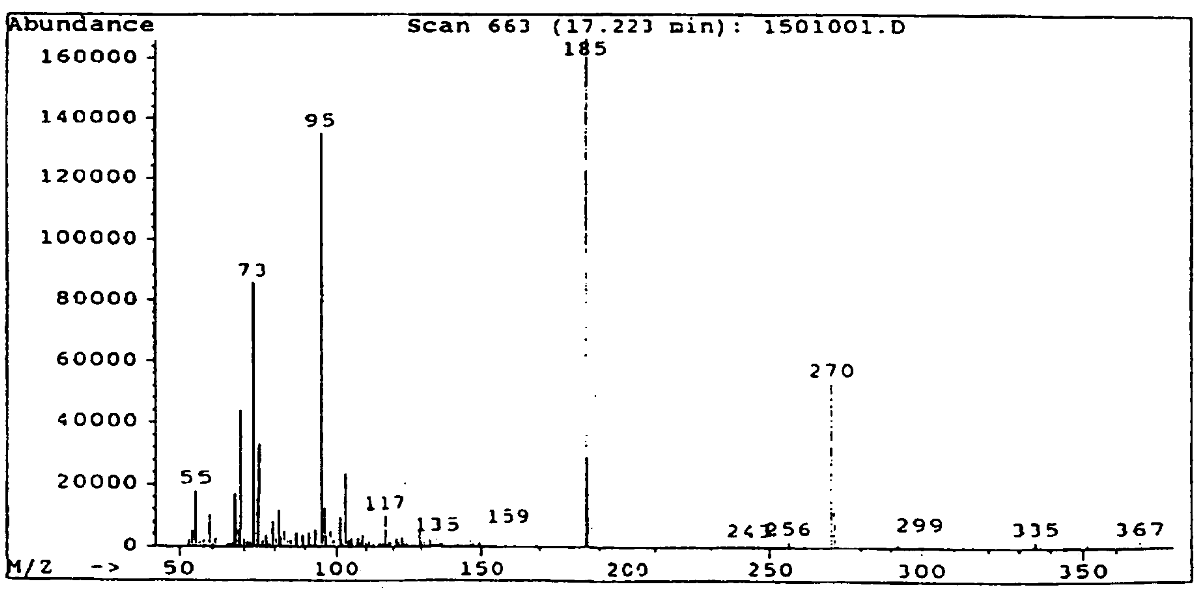

3A
Wild type
1
2
3
4
5
6
7
8
9
3B
Transgenic
1
2
3
4
5
6
7
10
11
8
12
9
13
10
15
20
Elution time (min)
Relative detector response

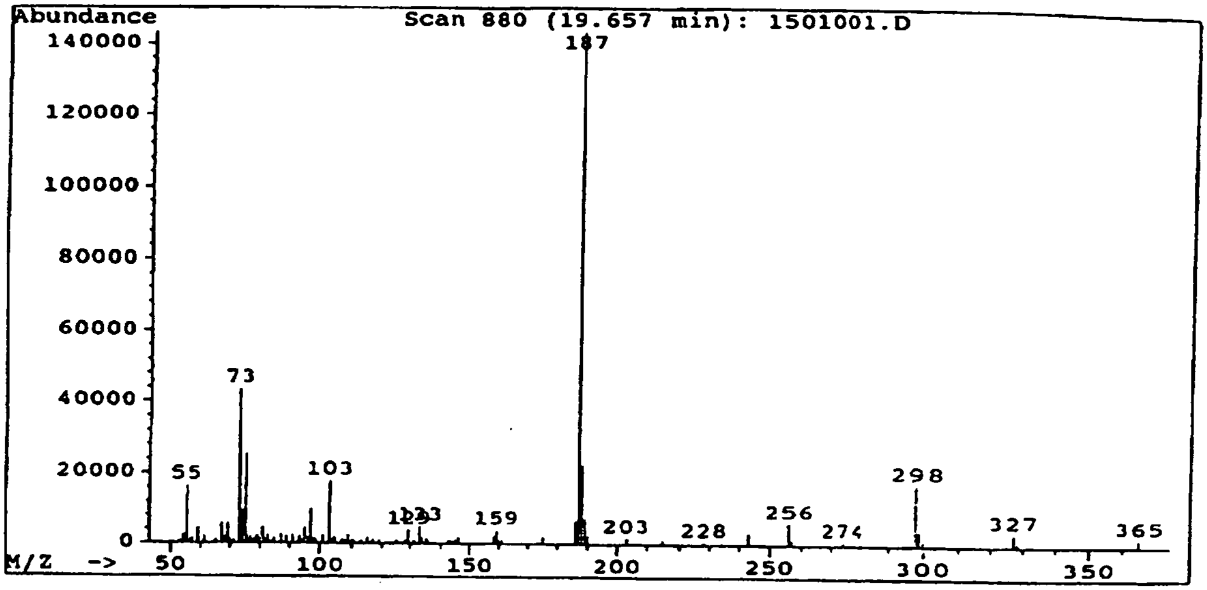
Figure 4A Mass spectrum of peak 10 from figure 3B



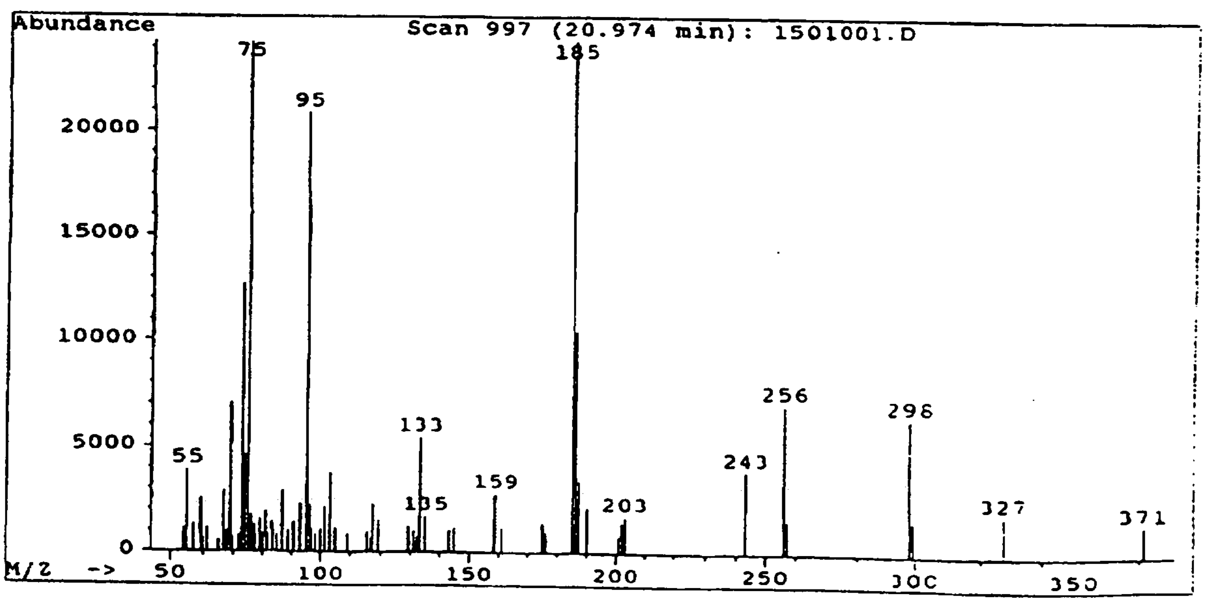
4B Mass spectrum of peak 11 from figure 3B



4C Mass spectrum of peak 12 from figure 3B



4D Mass spectrum of peak 13 from figure 3B



10	20	30	40	50	60
TATTGGCACC	GGCGGCACCA	TTCCAACAAT	GGATCCCTAG	AAAAAGATGA	AGTCTTTGTC
70	80	90	100	110	120
CCACCTAAGA	AAGCTGCAGT	CANATGGTAT	GTCAAATACC	TCAACAACCC	TCTTGGACGC
130	140	150	160	170	180
ATTCTGGTGT	TAACAGTTCA	GTTTATCCTC	GGGTGGCCTT	TGTATCTAGC	CTTTAATGTA
190	200	210	220	230	240
TCAGGTAGAC	CTTATGATGG	TTTCGCTTCA	CATTTCTTCC	CTCATGCACC	TATCTTTAAG
250	260	270	280	290	300
GACCGTGAAC	GTCTCCAGAT	ATACATCTCA	GATGCTGGTA	TTCTAGCTGT	CTGTTATGGT
310	320	330	340	350	360
CTTTACCGTT	ACGCTGCTTC	ACAAGGATTG	ACTGCTATGA	TCTGCGTCTA	CGGAGTACCG
370	380	390	400	410	420
CTTTTGATAG	TGAACTTTTT	CCTTGTCTTG	GTCACTTTCT	TGCAGCACAC	TCATCCTTCA
430	440	450	460	470	480
TTACCTCACT	ATGATTCAAC	CGAGTGGGAA	TGGATTAGAG	GAGCTTTGGT	TACGGTAGAC
490	500	510	520	530	540
AGAGACTATG	GAATCTTGAA	CAAGGTGTTT	CACAACATAA	CAGACACCCA	CGTAGCACAC
550					
CAC					

Figure 5

10	20	30	40	50	60
TATAGGCACC	GGAGGCACCA	TTCCAACACA	GGATCCCTCG	AAAGAGATGA	AGTATTTGTC
70	80	90	100	110	120
CCAAAGCAGA	AATCCGCAAT	CAAGTGGTAC	GGCGAATACC	TCAACAACCC	TCCTGGTCGC
130	140	150	160	170	180
ATCATGATGT	TAACTGTCCA	GTTCGTCCTC	GGATGGCCCT	TGTACTTAGC	CTTCAACGTT
190	200	210	220	230	240
TCTGGCAGAC	CCTACAATGG	TTTCGCTTCC	CATTTCTTCC	CCAATGCTCC	TATCTACAAC
250	260	270	280	290	300
GACCGTGAAC	GCCTCCAGAT	TTACATCTCT	GATGCTGGTA	TTCTAGCCGT	CTGTTATGGT
310	320	330	340	350	360
CTTTACCGTT	ACGCTGTTGC	ACAAGGACTA	GCCTCAATGA	TCTGTCTAAA	CGGAGTTCCG
370	380	390	400	410	420
CTTCTGATAG	TTAACTTTTT	CCTCGTCTTG	ATCACTTACT	TACAACACAC	TCACCCTGCG
430	440	450	460	470	480
TTGCCTCACT	ATGATTCATC	AGAGTGGGAT	TGGCTTAGAG	GAGCTTTAGC	TACTGTAGAC
490	500	510	520	530	540
AGAGACTATG	GAATCTTGAA	CAAGGTGTTC	CATAACATCA	CAGACACCCA	CGTCGCACAC
550					
CACT					

Figure 6

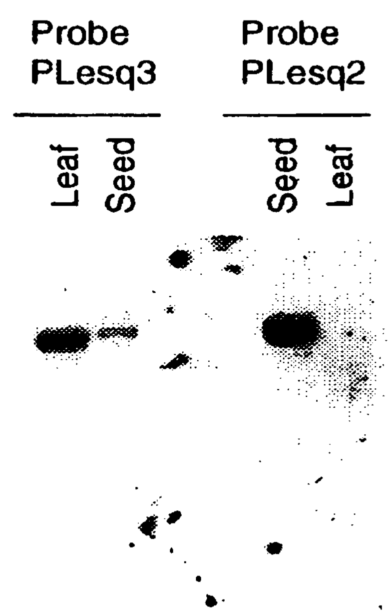


FIG.7

```
 AT GAA GCT TTA TAA GAA GTT AGT TTT CTC TGG TGA CAG AGA AAT TNT      47

GTC AAT TGG TAG TGA CAG TTG AAG CAA CAG GAA CAA CAA GGA TGG TTG      95

GTG NTG ATG CTG ATG TGG TGA TGT GTT ATT CAT CAA ATA CTA AAT ACT     143

ACA TTA CTT GTT GCT GCC TAC TTC TCC TAT TTC CTC CGC CAC CCA TTT     191

TGG ACC CAC GAN CCT TCC ATT TAA ACC CTC TCT CGT GCT ATT CAC CAG     239

AAG AGA AGC CAA GAG AGA GAG AGA GAG AAT GTT CTG AGG ATC ATT GTC     287

TTC TTC ATC GTT ATT AAC GTA AGT TTT TTT TGA CCA CTC ATA TCT AAA     335

ATC TAG TAC ATG CAA TAG ATT AAT GAC TGT TCC TTC TTT TGA TAT TTT     383

                        Met Gly Ala Gly Gly Arg Ile Met Val Thr      10
CAG CTT CTT GAA TTC AAG ATG GGT GCT GGT GGA AGA ATA ATG GTT ACC     431

Pro Ser Ser Lys Lys Ser Glu Thr Glu Ala Leu Lys Arg Gly Pro Cys      26
CCC TCT TCC AAG AAA TCA GAA ACT GAA GCC CTA AAA CGT GGA CCA TGT     479

Glu Lys Pro Pro Phe Thr Val Lys Asp Leu Lys Lys Ala Ile Pro Gln      42
GAG AAA CCA CCA TTC ACT GTT AAA GAT CTG AAG AAA GCA ATC CCA CAG     527

His Cys Phe Lys Arg Ser Ile Pro Arg Ser Phe Ser Tyr Leu Leu Thr      58
CAT TGT TTC AAG CGC TCT ATC CCT CGT TCT TTC TCC TAC CTT CTC ACA     575

Asp Ile Thr Leu Val Ser Cys Phe Tyr Tyr Val Ala Thr Asn Tyr Phe      74
GAT ATC ACT TTA GTT TCT TGC TTC TAC TAC GTT GCC ACA AAT TAC TTC     623

Ser Leu Leu Pro Gln Pro Leu Ser Thr Tyr Leu Ala Trp Pro Leu Tyr      90
TCT CTT CTT CCT CAG CCT CTC TCT ACT TAC CTA GCT TGG CCT CTC TAT     671

Trp Val Cys Gln Gly Cys Val Leu Thr Gly Ile Trp Val Ile Gly His     106
TGG GTA TGT CAA GGC TGT GTC TTA ACC GGT ATC TGG GTC ATT GGC CAT     719

Glu Cys Gly His His Ala Phe Ser Asp Tyr Gln Trp Val Asp Asp Thr     122
GAA TGT GGT CAC CAT GCA TTC AGT GAC TAT CAA TGG GTA GAT GAC ACT     767

Val Gly Phe Ile Phe His Ser Phe Leu Leu Val Pro Tyr Phe Ser Trp     138
GTT GGT TTT ATC TTC CAT TCC TTC CTT CTC GTC CCT TAC TTC TCC TGG     815

Lys Tyr Ser His Arg Arg His His Ser Asn Asn Gly Ser Leu Glu Lys     154
AAA TAC AGT CAT CGT CGT CAC CAT TCC AAC AAT GGA TCT CTC GAG AAA     863

Asp Glu Val Phe Val Pro Pro Lys Lys Ala Ala Val Lys Trp Tyr Val     170
GAT GAA GTC TTT GTC CCA CCG AAG AAA GCT GCA GTC AAA TGG TAT GTT     911

Lys Tyr Leu Asn Asn Pro Leu Gly Arg Ile Leu Val Leu Thr Val Gln     186
AAA TAC CTC AAC AAC CCT CTT GGA CGC ATT CTG GTG TTA ACA GTT CAG     959
```

Figure 8A

Phe Ile Leu Gly Trp Pro Leu Tyr Leu Ala Phe Asn Val Ser Gly Arg 202
TTT ATC CTC GGG TGG CCT TTG TAT CTA GCC TTT AAT GTA TCA GGT AGA 1007

Pro Tyr Asp Gly Phe Ala Ser His Phe Phe Pro His Ala Pro Ile Phe 218
CCT TAT GAT GGT TTC GCT TCA CAT TTC TTC CCT CAT GCA CCT ATC TTT 1055

Lys Asp Arg Glu Arg Leu Gln Ile Tyr Ile Ser Asp Ala Gly Ile Leu 234
AAA GAC CGA GAA CGC CTC CAG ATA TAC ATC TCA GAT GCT GGT ATT CTA 1103

Ala Val Cys Tyr Gly Leu Tyr Arg Tyr Ala Ala Ser Gln Gly Leu Thr 250
GCT GTC TGT TAT GGT CTT TAC CGT TAC GCT GCT TCA CAA GGA TTG ACT 1151

Ala Met Ile Cys Val Tyr Gly Val Pro Leu Leu Ile Val Asn Phe Phe 266
GCT ATG ATC TGC GTC TAT GGA GTA CCG CTT TTG ATA GTG AAC TTT TTC 1199

Leu Val Leu Val Thr Phe Leu Gln His Thr His Pro Ser Leu Pro His 282
CTT GTC TTG GTA ACT TTC TTG CAG CAC ACT CAT CCT TCG TTA CCT CAT 1247

Tyr Asp Ser Thr Glu Trp Glu Trp Ile Arg Gly Ala Leu Val Thr Val 298
TAT GAT TCA ACC GAG TGG GAA TGG ATT AGA GGA GCT TTG GTT ACG GTA 1295

Asp Arg Asp Tyr Gly Ile Leu Asn Lys Val Phe His Asn Ile Thr Asp 314
GAC AGA GAC TAT GGA ATA TTG AAC AAG GTG TTC CAT AAC ATA ACA GAC 1343

Thr His Val Ala His His Leu Phe Ala Thr Ile Pro His Tyr Asn Ala 330
ACA CAT GTG GCT CAT CAT CTC TTT GCA ACT ATA CCG CAT TAT AAC GCA 1391

Met Glu Ala Thr Glu Ala Ile Lys Pro Ile Leu Gly Asp Tyr Tyr His 346
ATG GAA GCT ACA GAG GCG ATA AAG CCA ATA CTT GGT GAT TAC TAC CAC 1439

Phe Asp Gly Thr Pro Trp Tyr Val Ala Met Tyr Arg Glu Ala Lys Glu 362
TTC GAT GGA ACA CCG TGG TAT GTG GCC ATG TAT AGG GAA GCA AAG GAG 1487

Cys Leu Tyr Val Glu Pro Asp Thr Glu Arg Gly Lys Lys Gly Val Tyr 378
TGT CTC TAT GTA GAA CCG GAT ACG GAA CGT GGG AAG AAA GGT GTC TAC 1535

Tyr Tyr Asn Asn Lys Leu 384
TAT TAC AAC AAT AAG TTA TGA GGC TGA TAG GGC GAG AGA AGT GCA ATT 1583

ATC AAT CTT CAT TTC CAT GTT TTA GGT GTC TTG TTT AAG AAG CTA TGC 1631

TTT GTT TCA ATA ATC TCA GAG TCC ATN TAG TTG TGT TCT GGT GCA TTT 1679

TGC CTA GTT ATG TGG TGT CGG AAG TTA GTG TTC AAA CTG CTT CCT GCT 1727

GTG CTG CCC AGT GAA GAA CAA GTT TAC GTG TTT AAA ATA CTC GGA ACG 1775

AAT TGA CCA CAA NAT ATC CAA AAC CGG CTA TCC GAA TTC CAT ATC CGA 1823

AAA CCG GAT ATC CAA ATT TCC AGA GTA CTT AG 1855

Figure 8B

```
                       10         20         30         40         50
LFFAH12     1 MGAGGRIM-- --VTPSSKKS --ETEALKRG PCEKPPFTVK DLKKAIPQHC    50
FAH12       1 MGGGGRMSTV ITSNNSEKKG --GSSHLKRA PHTKPPFTLG DLKRAIPPHC    50
ATFAD2      1 MGAGGRMP-- --VPTSSKKS --ETDTTKRV PCEKPPFSVG DLKKAIPPHC    50
BNFAD2      1 MGAGGRMQ-- --VSPPSKKS --ETDNIKRV PCETPPFTVG ELKKAIPPHC    50
GMFAD2-1    1 MGLA-KETTM GGRGRVAKVE VQGKKPLSRV PNTKPPFTVG QLKKAIPPHC    50
GMFAD2-2    1 MGAGGR---- TDVPPANRKS --EVDPLKRV PFEKPQFSLS QIKKAIPPHC    50
ZMFAD2      1 MGAGGRMTEK EREKQEQLAR ATGGAAMQRS PVEKPPFTLG QIKKAIPPHC    50
RCFAD2      1 ---------- ---------- ---------- ---------- ----------    50
                       60         70         80         90        100
LFFAH12    51 FKRSIPRSFS YLLTDITLVS CFYYVATNYF SLLPQPLSTY LAWPLYWVCQ   100
FAH12      51 FERSFVRSFS YVAYDVCLSF LFYSIATNFF PYISSPLS-Y VAWLVYWLFQ   100
ATFAD2     51 FKRSIPRSFS YLISDIIIAS CFYYVATNYF SLLPQPLS-Y LAWPLYWACQ   100
BNFAD2     51 FKRSIPRSFS HLIWDIIIAS CFYYVATTYF PLLPNPLS-Y FAWPLYWACQ   100
GMFAD2-1   51 FQRSLLTSFS YVVYDLSFAF IFY-IATTYF HLLPQPFS-L IAWPIYWVLQ   100
GMFAD2-2   51 FQRSVLRSFS YVVYDLTIAF CLYYVATHYF HLLPGPLS-F RGMAIYWAVQ   100
ZMFAD2     51 FERSVLKSFS YVVHDLVIAA ALLYFALAII PALPSPLR-Y AAWPLYWIAQ   100
RCFAD2     51 ---------- ---------- ---------- ---------- ----------   100
                      110        120        130        140        150
LFFAH12   101 GCVLTGIWVI GHECGHHAFS DYQWVDDTVG FIFHSFLLVP YFSWKYSHRR   150
FAH12     101 GCILTGLWVI GHECGHHAFS EYQLADDIVG LIVHSALLVP YFSWKYSHRR   150
ATFAD2    101 GCVLTGIWVI AHECGHHAFS DYQWLDDTVG LIFHSFLLVP YFSWKYSHRR   150
BNFAD2    101 GCVLTGVWVI AHECGHAAFS DYQWLDDTVG LIFHSFLLVP YFSWKYSHRR   150
GMFAD2-1  101 GCLLTGVWVI AHECGHHAFS KYQWVDDVVG LTLHSTLLVP YFSWKISHRR   150
GMFAD2-2  101 GCILTGVWVI AHECGHHAFS DYQLLDDIVG LILHSALLVP YFSWKYSHRR   150
ZMFAD2    101 G--------- -------AFS DYSLLDDVVG LVLHSSLMVP YFSWKYSHRR   150
RCFAD2    101 -------WVM AHDCGHHAFS DYQLLDDVVG LILHSCLLVP YFSWKHSHRR   150
                      160        170        180        190        200
LFFAH12   151 HHSNNGSLEK DEVFVPPKKA AVKWYVKYL- NNPLGRILVL TVQFILGWPL   200
FAH12     151 HHSNIGSLER DEVFVPKSKS KISWYSKYS- NNPPGRVLTL AATLLLGWPL   200
ATFAD2    151 HHSNTGSLER DEVFVPKQKS AIKWYGKYL- NNPLGRIMML TVQFVLGWPL   200
BNFAD2    151 HHSNTGSLER DEVFVPR-RS QTSSGTAST- STTFGRTVML TVQFTLGWPL   200
GMFAD2-1  151 HHSNTGSLDR DEVFVPKPKS KVAWFSKYL- NNPLGRAVSL LVTLTIGWPM   200
GMFAD2-2  151 HHSNTGSLER DEVFVPKQKS CIKWYSKYL- NNPPGRVLTL AVTLTLGWPL   200
ZMFAD2    151 HHSNTGSLER DEVFVPKKKE ALPWYTPYVY NNPVGRVVHI VVQLTLGWPL   200
RCFAD2    151 HHSNTGSLER DEVFVPKKKS SIRWYSKYL- NNPPGRIMTI AVTLSLGWPL   200
                      210        220        230        240        250
LFFAH12   201 YLAFNVSGRP YDG-FASHFF PHAPIFKDRE RLQIYISDAG ILAVCYGLYR   250
FAH12     201 YLAFNVSGRP YDR-FACHYD PYGPIFSERE RLQIYIADLG IFATTFVLYQ   250
ATFAD2    201 YLAFNVSGRP YDG-FACHFF PNAPIYNDRE RLQIYLSDAG ILAVCFGLYR   250
BNFAD2    201 YLAFNVSGRP YDGGFACHFH PNAPIYNDRE RLQIYISDAG ILAVCYGLLP   250
GMFAD2-1  201 YLAFNVSGRP YDS-FASHYH PYAPIYSNRE RLLIYVSDVA LFSVTYSLYR   250
GMFAD2-2  201 YLALNVSGRP YDR-FACHYD PYGPIYSDRE RLQIYISDAG VLAVVYGLFR   250
ZMFAD2    201 YLATNASGRP YPR-FACHFD PYGPIYNDRE RAQIFVSDAG VVAVAFGLYK   250
RCFAD2    201 YLAFNVSGRP YDR-FACHYD PYGPIYNDRE RIEIFISDAG VLAVTFGLYQ   250
```

Figure 9A

```
                          260        270        280        290        300
LFFAH12   251  YAASQGLTAM ICVYGVPLLI VNFFLVLVTF LQHTHPSLPH YDSTEWEWIR      300
FAH12     251  ATMAKGLAWV MRIYGVPLLI VNCFLVMITY LQHTHPAIPR YGSSEWDWLR      300
ATFAD2    251  YAAAQGMASM ICLYGVPLLI VNAFLVLITY LQHTHPSLPH YDSSEWDWLR      300
BNFAD2    251  YAAVQGVASM VCFLRVPLLI VNGFLVLITY LQHTHPSLPH YDSSEWDWLR      300
GMFAD2-1  251  VATLKGLVWL LCVYGVPLLI VNGFLVTITY LQHTHFALPH YDSSEWDWLK      300
GMFAD2-2  251  LAMAKGLAWV VCVYGVPLLV VNGFLVLITF LQHTHPALPH YTSSEWDWLR      300
ZMFAD2    251  LAAAFGVWWV VRVYAVPLLI VNAWLVLITY LQHTHPSLPH YDSSEWDWLR      300
RCFAD2    251  LAIAKGLAWV VCVYGVPLLV VNSFLVLITF LQHTHPALPH YDSSEWDWLR      300

                          310        320        330        340        350
LFFAH12   301  GALVTVDRDY GILNKVFHNI TDTHVAHHLF ATIPHYNAME ATEAIKPILG      350
FAH12     301  GAMVTVDRDY GVLNKVFHNI ADTHVAHHLF ATVPHYHAME ATKAIKPIMG      350
ATFAD2    301  GALATVDRDY GILNKVFHNI TDTHVAHHLF STMPHYNAME ATKAIKPILG      350
BNFAD2    301  GALATVDRDY GILNQGFHNI TDTHEAHHLF STMPHYHAME ATKAIKPILG      350
GMFAD2-1  301  GALATMDRDY GILNKVFHHI TDTHVAHHLF STMPHYHAME ATNAIKPILG      350
GMFAD2-2  301  GALATVDRDY GILNKVFHNI TDTHVAHHLF STMPHYHAME ATKAIKPILG      350
ZMFAD2    301  GALATMDRDY GILNRVFHNI TDTHVAHHLF STMPHYHAME ATKAIRPILG      350
RCFAD2    301  GALATVDRDY GILNKVFHNI TDTQVAHHLF ---------- ----------      350
                          360        370        380        390        400
LFFAH12   351  DYYHFDGTPW YVAMYREAKE CLYVEPDTER GKKGVYYYNN K-L.......      400
FAH12     351  EYYRYDGTPF YKALWREAKE CLFVEPDEGA PTQGVFWYRN KY-.......      400
ATFAD2    351  DYYQFDGTPW YVAMYREAKE CIYVEPDREG DKKGVYWYNN K-L.......      400
BNFAD2    351  EYYQFDGTPV VKAMWREAKE CIYVEPDRQG EKKGVFWYNN KL*.......      400
GMFAD2-1  351  EYYQFDDTPF YKALWREARE CLYVEPDEGT SEKGVYWYRN KY-.......      400
GMFAD2-2  351  EYYRFDETPF VKAMWREARE CIYVEPDQST ESKGVFWYNN KL-.......      400
ZMFAD2    351  DYYHFDPTPV AKATWREAGE CIYVEPE--- DRKGVFWYNK KF*.......      400
```

Figure 9B

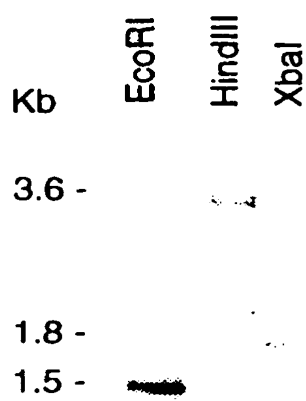
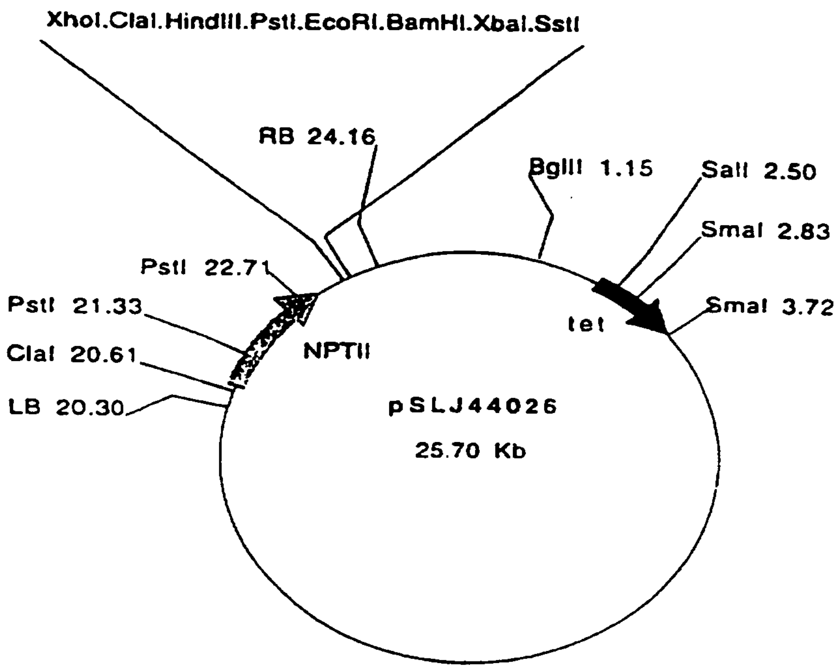


FIG.10



Plasmid name: pSLJ44026
Plasmid size: 25.70 kb
Constructed by: Jonathon Jones
Construction date: 1992
Comments/References: Transgenic Research 1,285-297 (1992)

Figure 11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/02187

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :A01H 5/00, 5/10; C12N 15/52; 15/82
US CL :800/205; 435/172.3, 419; 536/23.6
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 800/205; 435/172.3, 419; 536/23.6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ---- Y	WO 94/11516 A1 (LIGHTNER et al) 26 May 1994, especially pages 40-44 and 109.	1, 3, 7, 10, 11, 16, 17, 22, 23, 27-34 ---------------- 2, 4-9, 12-15, 18, 21, 24-26

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
03 JUNE 1997	30 JUL 1997

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	ELIZABETH F. MCELWAIN
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

13

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/82, 15/53, 15/11, 5/10, A01H 5/00, C11B 1/00 — A1

(11) International Publication Number: WO 94/18337

(43) International Publication Date: 18 August 1994 (18.08.94)

(21) International Application Number: PCT/US94/01321

(22) International Filing Date: 4 February 1994 (04.02.94)

(30) Priority Data:
014,431 5 February 1993 (05.02.93) US
156,551 22 November 1993 (22.11.93) US

(60) Parent Applications or Grants
(63) Related by Continuation
US 014,431 (CIP)
Filed on 5 February 1993 (05.02.93)
US 156,551 (CIP)
Filed on 22 November 1993 (22.11.93)

(71) Applicants *(for all designated States except US)*: MONSANTO COMPANY [US/US]; 800 North Lindbergh Boulevard, St. Louis, MO 63167 (US). MICHIGAN STATE UNIVERSITY [US/US]; 238 Administration Building, East Lansing, MI 48824 (US).

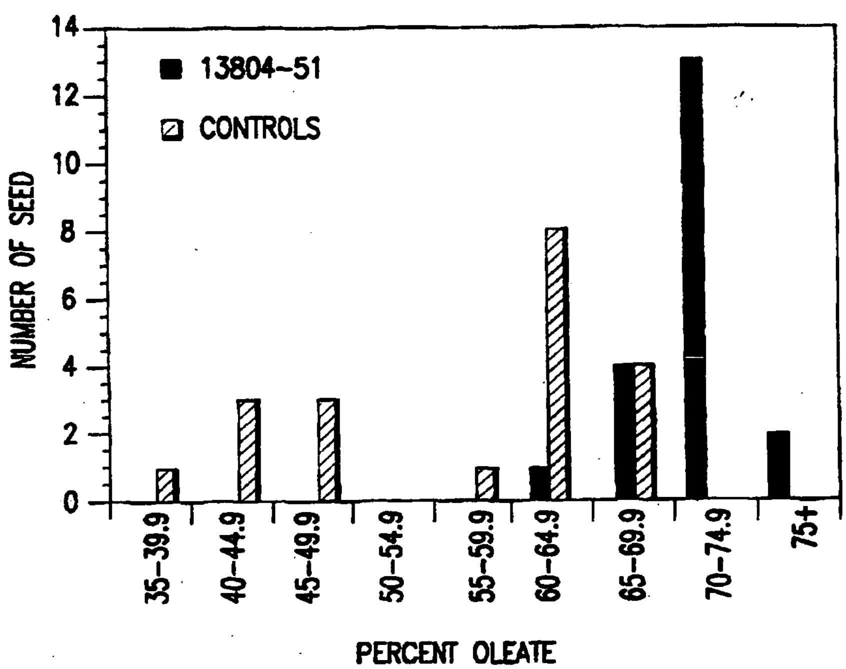
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: GIBSON, Susan, Irma [US/US]; Department of Biochemistry and Cellular Biology, Rice University, Houston, TX 77251 (US). KISHORE, Ganesh, Murthy [US/US]; 15354 Grantley Drive, Chesterfield, MO 63017 (US). RUFF, Thomas, Gene [US/US]; 5500 Sassafras Lane, High Ridge, MO 63049 (US). SOMERVILLE, Christopher, Roland [CA/US]; 480 Hale Street, Palo Alto, CA 94301 (US). ARONDEL, Vincent, Jean-Marie, Armel [FR/FR]; Laboratoire de Physiologie Cellulaire et Moléculaire, Tour 53, 3ème Etage, Université Pierre-et-Marie-Curie, 4, place Jussieu, F-75252 Paris Cédex 05 (FR).

(74) Agent: BOLDING, James, Clifton; Monsanto Company, 800 North Lindberg Boulevard, St. Louis, MO 63167 (US).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ALTERED LINOLENIC AND LINOLEIC ACID CONTENT IN PLANTS



(57) Abstract

Transformed plants which have increased or decreased linolenic acid content are disclosed. Also disclosed are plants which express a linoleic acid desaturase gene.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SI	Slovenia
SK	Slovakia
SN	Senegal
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

ALTERED LINOLENIC AND LINOLEIC ACID CONTENT IN PLANTS

This is a continuation-in-part of U.S. Serial No. 08/156,551 filed November 22, 1993, which is a continuation of U.S. Serial No. 08/014,431, filed on February 5, 1993. The present invention relates to genetically engineered plants. In particular it relates to genetically engineered plants and seeds which have altered linolenic and linoleic acid content compared with naturally occurring plants.

BACKGROUND

Many crop species produce seed oils in which the fatty acid composition is not ideally suited to the intended use. The application of conventional breeding methods, coupled in some cases with mutagenesis, has resulted in the production of new varieties of several species with desirable alterations in the fatty acid composition of seed oil. A notable example is the development of low erucic acid varieties of rapeseed (Stefansson 1983). Similar efforts have resulted in the reduction of the level of polyunsaturated 18-carbon fatty acids in soybean (Wilcox and Cavins 1985; Graef et al. 1988), sunflower (Fick 1989), and linseed oils (Green and Marshal 1984).

Most of the genetic variation in seed lipid fatty acid composition appears to involve the presence of an allele of a gene that disrupts normal fatty acid metabolism and leads to an accumulation of intermediate fatty acid products in the seed storage lipids (Downey 1987). However, it seems likely that, because of the inherent limitations of this approach, many other desirable changes in seed oil fatty acid composition may require the directed application of genetic engineering methods.

α-Linolenic acid ($18:3^{\Delta 9,12,15}$) is an eighteen carbon fatty acid containing three *cis* double bonds at the 9-10, 12-13 and 15-16 carbons. It is found in the cells of higher plants as a constituent of cell membranes. It

is also found in storage organs, such as in seeds. There it is designated oil bodies which are bounded by an electron dense structure that is thought to be a half-unit membrane and dispersed in the cytoplasmic environment of cells. When present as a constituent of cell membranes, linolenic acid is usually esterified to the sn-1 or sn-2 position of the glycerol moiety of a diacyl-glycerolipid. By contrast, when present in oil bodies, linolenic acid is usually esterified to the sn-1, sn-2 or sn-3 position of a triacylglycerolipid (TAG).

Linolenic acid is extensively used in the paint and varnish industry in view of its rapid oxidation. Flax seed is a predominant source of this oil. Soybean seed, on the other hand, does not have sufficient linolenic acid content to be used in this industry. Thus, increasing the linolenic acid content in a plant such as soybean would permit the use of the soybean oil in the paint and varnish industry.

On the other hand, it is undesirable to have significant levels of linolenic acid in cooking oils and foods. Linolenic acid is unstable during cooking and is rapidly oxidized. The oxidized products impart rancidity to the finished product. A rapeseed or soybean oil with reduced linolenic acid, such as containing 2% or less of linolenic acid, would be ideal for use as a cooking oil.

Linolenic acid is also a precursor in the biosynthesis of jasmonic acid, an important plant growth regulator. Linolenic acid is converted to jasmonic acid by introduction of an oxygen to the carbon chain by a lipoxygenase, followed by dehydration, reduction, and several β-oxidations (Vick and Zimmerman, 1984). The activity of jasmonic acid has been measured in terms of induction of pathogen defense responses. By application of free linolenic acid to plants, plant pathogen defenses can also be induced (Farmer and Ryan, 1992).

A model has been proposed to explain the ability of free linolenic acid to exhibit the effects associated with jasmonic acid (Farmer and Ryan, 1992). It is hypothesized that all of the enzymatic activities which are required for the conversion of linolenic acid to jasmonic acid are constitutively present in the cell and the rate limiting step in the production of jasmonic acid is the availability of free linolenic acid. A likely route for the production of the free linolenic acid is by the activity of a lipase in the plasma membrane.

It has been observed that exogenous jasmonic acid can more powerfully activate defense responses than can wounding. This suggests that wounds cannot generate enough free linolenic acid to support high level production of jasmonic acid. The activity of the lipase or the availability of appropriate substrate for the lipase may be rate limiting upon wounding. Thus, increasing the linolenic acid content of plasma membrane may positively influence "signal transduction" in plants and result in better protection against environment and pathogen stress.

Linolenic acid, as well as oleic and linoleic acids are also important constituents, as well as precursors of volatile carbonyl compounds, whic contribute to the aroma of both fresh and cooked foods. The major fatty acids of tomato fruit pericarp are oleic, linoleic and linolenic acids. As the fruit ripens, the levels of the latter two fatty acids decline resulting in the production of a number of 4-6 carbon containing aldehydees and ketones. One particular metabolite, *cis*-3-hexanol, has been shown to be present in higher levels in vine-ripened tomatoes compared to supermarket tomatoes or tomatoes stored in refrigerators. It is likely, therefore, that the "aroma" of fresh fruits and vegetables can be "modulated" by regulation of the content of linolenic and linoleic acids, important substrates for the enzyme lipoxygenase and subsequently the

hydroperoxide cleaving enzyme, which generates the volatile "aroma" compounds.

From the above, it is clear that the ability to vary the content of linolenic acid in plants would be desirable. However, to achieve this result it is necessary to determine what controls the product of linolenic acid in plants.

A large body of experimental evidence derived from radiochemical tracer studies has indicated that α-linolenic acid is synthesized by the desaturation of linoleic acid ($18{:}2^{\Delta 9,12}$) (reviewed in Harwood 1988;). However, the actual substrate for desaturation is not known.

In vivo and *in vitro* labelling studies suggest that there are possibly two distinct pathways for the synthesis of linolenic acid (Browse and Somerville, 1991). One possible pathway is thought to be located in the endoplasmic reticulum where linoleic acid esterified to the sn-2 position of phosphatidylcholine is a substrate for desaturation. However, the available evidence does not exclude the possibility that linoleic acid esterified to other lipids may also be a substrate.

A second possible pathway of linoleic acid desaturation is located in the plastid where the available evidence suggests that linoleic acid esterified to monogalactosyldiacylglycerol and, possibly, other plastid lipids is the substrate for desaturation.

Relatively little direct information is available concerning the enzymes involved in linoleic acid desaturation. Low levels of enzyme activity have been detected in microsomal membrane preparations from developing linseed (Linum ussitatum) (Browse and Slack, 1981) and, more recently, in preparations of gently lysed chloroplasts (Schmidt and Heinz, 1990a,b). The general features of the enzyme may be inferred from information available about other enzymes of this class.

The most thoroughly characterized desaturase is the stearoyl-Coenzyme A (CoA) desaturase from vertebrate liver (reviewed by Holloway, 1983). This enzyme has been shown to be an integral membrane protein which contains non-heme iron. The desaturase reaction requires fatty acyl-CoA, molecular oxygen and reduced cytochrome b5, another membrane protein. *In vivo*, the reduced cytochrome b5 is produced by the transfer of reducing equivalents from NADH via the activity of cytochrome b5 reductase, a flavin containing membrane protein.

The most thoroughly characterized desaturase from plants is the stearoyl-ACP desaturase (McKeon and Stumpf, 1982; Shanklin and Somerville, 1991). This enzyme also requires molecular oxygen and a high potential reductant. However, in contrast to the animal enzyme, this desaturase is a soluble plastid protein which preferentially acts on a fatty acid esterified to acyl carrier protein (ACP) rather than CoA. This enzyme also differs from the animal enzyme by utilizing reduced ferredoxin as an intermediate electron donor.

Other plant desaturases appear to be membrane proteins. The microsomal Δ12 oleate desaturase from several plant species has been assayed in membrane preparations from several plants (Harwood, 1988). As with the stearoyl-CoA desaturase from animals, this enzyme requires molecular oxygen and reduced cytochrome b5 as an electron donor (Kearns et al., 1991). However, it appears that oleate esterified to a phospholipid is the substrate rather than a CoA ester.

With regard to the activity responsible for the making of linolenic acid, little was known as to its source or origin. However, evidence that the amount of linolenic acid is related to the amount of linoleic acid desaturase activity has been obtained by analysis of the properties of the fad3 mutant of *Arabidopsis thaliana* (Lemieux et al. 1990). This mutant is deficient in linolenic acid in the storage oils of its seed lipids and in the

membrane lipids of different tissues to varying degrees. The mutant also had an increase in the amount of linoleic acid. This can be interpreted as evidence that the mutant is defective in the activity of a desaturase which converts linoleic acid to linolenic acid.

There is further evidence to suggest that the activity of this desaturase could be rate limiting for linolenic acid synthesis under normal circumstances. This was discovered by measuring the effects on fatty acid composition in heterozygous plants (i.e., fad3+/fad-) formed by crossing the wild type with the fad3 mutant. In these F1 plants, which have one copy of the normal fad3 gene product instead of the two normally found in the wild type, the amount of linolenic acid was almost exactly intermediate between that found in either parent. This suggests that the amount of linolenic acid is proportional to the amount of functional fad3 gene product (Lemieux et al., 1990).

These results do not shed any light, however, on the nature of the fad3 gene product or whether the observed effects in mutants are related to either a decrease in quantitiy of desaturase protein or desaturase activity due to a defective protein.

Moreover, nothing is known with any degree of certainty about the linoleic acid desaturase from plant microsomes. As noted above, very little is known about the microsomal desaturases except that they probably utilize reduced cytochrome b5 as intermediate electron donor and probably utilize lipids rather than CoA or ACP esters as substrates.

Moreover, as in many other aspects of plant biology, the lack of specific information about the biochemistry and regulation of lipid metabolism makes it difficult to predict how the introduction of one or a few genes might usefully alter seed lipid synthesis.

An additional problem arises from the fact that many of the key enzymes of lipid metabolism are membrane-bound and present in low

quantities. Thus, attempts to solubilize and purify them from plant sources have not been successful.

SUMMARY OF THE INVENTION

The present invention provides structural coding sequences encoding linoleic acid desaturase activity which can be used to alter the linoleic and linolenic acid compositions of plants or to isolate other plant linoleic acid desaturases. The present invention further provides a plant capable of expressing a structural coding sequence to control the level of linolenic acid or linoleic acid or both in the plant. The present invention further provides a method for controlling the levels of linoleic and linolenic acid in plants. It is also demonstrated by the present invention that the linoleic acid desaturase enzyme activity in plant cells and tissues is a controlling step in linolenic acid biosynthesis.

The present invention further relates to the engineering of two advantageous traits into plants: increased and decreased α–linolenic acid content in the structural lipids or storage oils of various crop plants.

In accomplishing the foregoing, there is provided, in accordance with one aspect of the present invention, a genetically transformed plant which has an elevated linolenic acid content comprising a recombinant, double-stranded DNA molecule comprising

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a structural coding sequence that causes the production of an RNA sequence that encodes a linoleic acid desaturase activity; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

In accordance with another aspect of the present invention, there is provided a genetically transformed plant which has a reduced linolenic acid content, comprising a recombinant, double-stranded DNA molecule comprising

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a DNA sequence that causes the production of an RNA sequence that is in antisense orientation to at least a portion of a gene that encodes a linoleic acid desaturase activity in said plant; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

There has also been provided, in accordance with another aspect of the present invention a method of producing a genetically transformed plant which has an elevated or reduced linolenic acid content. There has also been provided, in accordance with another aspect of the present invention a recombinant, double-stranded DNA molecule and plant cells containing a recombinant, double-stranded DNA molecule.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the genetic map of the region of chromosome 2 of *Arabidopsis thaliana* where a linoleic acid desaturase gene is located and the identity of the yeast artificial chromosomes which carry this region of the genome.

Figure 2 shows the structure of plasmid pBNDES3 which was obtained by inserting an EcoRI fragment containing the *B. napus* linoleic acid desaturase cDNA (fad3) into pBLUESCRIPT.

Figure 3 shows the nucleotide sequence (SEQ ID NO:1) and deduced amino acid sequence (SEQ ID NO:2) for the linoleic acid desaturase cDNA (fad3) from *B. napus*.

Figure 4 shows a comparison of the deduced amino acid sequence of one linoleic acid desaturase cDNA (fad3) from *B. napus* and the desA gene from *Synechocystis*. Identical residues are indicated by a solid box. Conservative substitutions are indicated by a stippled box.

Figure 5 shows the binary Ti plasmid vector pBI121.

Figure 6 shows the binary Ti plasmid pTiDES3 which was constructed by insertion of a linoleic acid desaturase cDNA (fad3) into pBI121.

Figure 7 shows the map of the plant transformation vector pMON13804.

Figure 8 shows the map of the plant transformation vector pMON13805.

Figure 9 shows the oil content of control and transformed canola seed in accordance with the present invention.

Figure 10 shows the nucleotide sequence (SEQ ID NO:9) for the linoleic acid desaturase cDNA (fadD) from *Arabidopsis*.

Figure 11 shows the deduced amino acid sequence (SEQ ID NO:10) for the linoleic acid desaturase cDNA (fadD) from *Arabidopsis*.

Figure 12 shows the nucleotide sequence (SEQ ID NO:11) for the linoleic acid desaturase cDNA (fadE) from *Arabidopsis*.

Figure 13 shows the deduced amino acid sequence (SEQ ID NO:12) for the linoleic acid desaturase cDNA (fadE) from *Arabidopsis*.

DETAILED DESCRIPTION OF THE INVENTION

A genetically transformed plant of the present invention which has an altered linolenic or linoleic acid content can be obtained by

expressing the double-stranded DNA molecules described in this application.

The expression of a double-stranded DNA involves transcription of messenger RNA (mRNA) from one strand of the DNA by RNA polymerase enzyme, and the subsequent processing of the mRNA primary transcript inside the nucleus. This processing involves a 3' non-translated region which adds polyadenylate nucleotides to the 3' end of the RNA.

Promoters

Transcription of DNA into mRNA is regulated by a region of DNA usually referred to as the "promoter." The promoter region contains a sequence of bases that signals RNA polymerase to associate with the DNA, and to initiate the transcription of mRNA using one of the DNA strands as a template to make a corresponding complementary strand of RNA.

Any promoter which is known or is found to cause transcription of RNA in plant cells can be used in the present invention. Promoters which are useful in the present invention include any promoter that functions in a plant cell to cause the production of a RNA sequence. A number of promoters which are active in plant cells and are capable of producing a RNA sequence have been described in the literature. These include the nopaline synthase (NOS) and octopine synthase (OCS) promoters (which are carried on tumor-inducing plasmids of Agrobacterium tumefaciens), the caulimovirus promoters such as the cauliflower mosaic virus (CaMV) 19S and 35S and the figwort mosaic virus 35S-promoters, the light-inducible promoter from the small subunit of ribulose-1,5-bis-phosphate carboxylase (ssRUBISCO, a very abundant plant polypeptide), and the chlorophyll a/b binding protein gene promoter, etc. All of these promoters have been used to create various types of DNA constructs

which have been expressed in plants; see, e.g., PCT publication WO 84/02913 (Rogers et al., Monsanto).

Promoters may be obtained from a variety of sources such as plants and plant viruses. Promoters can be used in the form that they exist as isolated from plant genes such as ssRUBISCO genes, or can be modified to improve their effectiveness, such as with the enhanced CaMV35S promoter.

Those skilled in the art will recognize that the amount of linoleic acid desaturase needed to induce the desired alteration in linolenic acid content may vary with the type of plant. It is also possible that extremes in linoleic acid desaturase activity may be deleterious to the plant. Therefore, in a preferred embodiment, promoter function should be optimized by selecting a promoter with the desired tissue expression capabilities and approximate promoter strength and selecting a transformant which produces the desired linoleic acid desaturase activity in the target tissues.

This selection approach from the pool of transformants is routinely employed in expression of heterologous structural genes in plants since there is variation between transformants containing the same heterologous gene due to the site of gene insertion within the plant genome. (Commonly referred to as "position effect").

In a preferred embodiment, the promoters utilized in the double-stranded DNA molecules should have relatively high expression in tissues where the increased or decreased linolenic acid content is desired, such as the seeds of the plant. In Canola, a particularly preferred promoter in this regard is the seed specific promoter described herein in greater detail in the accompanying examples.

In another preferred embodiment, the promoter used in the expression of the double-stranded DNA molecules of the present invention

can be a constitutive promoter, expressing the DNA molecule in all or most of the tissues of the plant. However, the promoter selected for this embodiments should not cause expression at levels which are detrimental to plant health, growth and development.

ß-conglycinin (also known as the 7S protein) is one of the major storage proteins in soybean (Glycine max) (Meinke et al., 1981). The 7S (β-conglycin) α'-subunit promoter, used in one aspect of this study to express the linoleic acid desaturase gene, has been shown to be both highly active and seed-specific (Doyle et al, 1986 and Beachy et al., 1985). The ß-subunit of ß-conglycinin has been expressed, using its endogenous promoter, in the seeds of transgenic petunia and tobacco, showing that the promoter functions in a seed-specific manner in other plants (Bray et al., 1987). The promoter for ß-conglycinin could be used to in accordance with the present invention. If used, this promoter could express the DNA molecule specifically in seeds, which could lead to an alteration in the linolenic acid content of the seeds.

In addition, the endogenous plant linoleic acid desaturase promoters can be used in the present invention. These promoters should be useful in expressing a linoleic acid desaturase gene in specific tissues, such as leaves, seeds or fruits. A number of other promoters with seed-specific or seed-enhanced expression are known and are likely to be expressed in seeds, which are oil accumulating cells. For illustration, the napin promoter and the acyl carrier protein promoters have been utilized in the modification of seed oil by antisense expression (Knutson et al., 1992).

The linolenic acid content of root tissue can be increased by expressing a linoleic acid desaturase gene behind a promoter which is expressed in roots. The promoter from the acid chitinase gene (Samac et al., 1990) is known to function in root tissue and could be used to express the linoleic acid desaturase in root tissue. Expression in root tissue could

also be accomplished by utilizing the root specific subdomains of the CaMV35S promoter that have been identified. (Benfey et al., 1989). The linolenic acid content of leaf tissue can be increased by expressing the linoleic acid desaturase gene using a leaf active promoter such as ssRUBISCO promoter or chlorophyll a/b binding protein gene promoter.

The linolenic acid content of fruits can be increased by expressing a linolenic acid desaturase gene behind a promoter which is functional in fruits. Such promoters could be either expressed at all developmental stages of the fruit or restricted to specific stages, particularly fruit ripening.

The RNA produced by a DNA construct of the present invention can also contain a 5' non-translated leader sequence. This sequence can be derived from the promoter selected to express the gene, and can be specifically modified so as to increase translation of the mRNA. The 5' non-translated regions can also be obtained from viral RNAs, from suitable eukaryotic genes, or from a synthetic gene sequence. The present invention is not limited to constructs, as presented in the following examples, wherein the non-translated region is derived from the 5' non-translated sequence that accompanies the promoter sequence. Rather, the non-translated leader sequence can be derived from an unrelated promoter or coding sequence as discussed above.

Linoleic Acid Desaturase Structural Coding Sequences

The structural coding sequence that causes the production of an RNA sequence that encodes a linoleic acid desaturase activity can be the sequences disclosed in the present application, or any sequence that can be obtained using the sequences disclosed in the present application, or any sequence that can be isolated using the method disclosed in the present application.

The structural coding sequence can also be a part of or from the structural coding sequences disclosed in the present invention. It is possible that the active part of the linoleic acid desaturase is formed using only part of the structural coding sequences disclosed in the present application.

The structural coding sequences can be obtained from a variety of sources, such as algae, bacteria or plants. Preferably, structural coding sequences obtained from plants are used in accordance with the present invention.

Since virtually nothing was known about the properties of the linoleic acid desaturase structural coding sequence prior to the present invention, the method used in the present invention to isolate the structural coding sequence was based on the concept of map based cloning. The essential concept in map based cloning is to use information about the genetic map position of a structural coding sequence to isolate the region of the chromosome surrounding the structural coding sequence, and then to use the isolated DNA to complement a mutation in the structural coding sequence. This strategy has never previously been reported in the isolation of any plant gene.

In order to implement map based cloning of the linoleic acid desaturase, mutants of *Arabidopsis thaliana* (L.) deficient in linoleic acid desaturase activity were isolated by screening randomly chosen individuals from mutagenized populations of plants for individual plants with altered leaf or seed fatty acid composition. (Browse et al. 1985; Lemieux et al. 1990). By screening thousands of plants for altered fatty acid composition, mutants with decreased amounts of linolenic acid and increased amounts of linoleic acid in leaf and seed lipids were isolated. Physiological and genetic analyses of these mutants indicated that they fell into three complementation groups designated fad3, fadD and fadE.

The fad3 mutants had very reduced levels of linolenic acid in seeds and roots but had almost normal levels of linolenic acid in leaves. This effect was interpreted as evidence that the fad3 locus encoded a microsomal desaturase which was responsible for desaturation of linoleic acid to linolenic acid on lipids made by the pathway of lipid biosynthesis in the endoplasmic reticulum, designated the "eukaryotic pathway" (Lemieux et al. 1990). This pathway is mostly responsible for the synthesis of lipids in non-green tissues such as seeds and roots, but plays a secondary role in leaves and other green tissues. Thus, a mutation in the fad3 gene would not be expected to have a major effect on the desaturation of leaf lipids.

In contrast to the fad3 mutant, the fadD mutant had almost normal fatty acid composition of roots and seeds, but had a strong reduction in the amount of linolenic acid in leaf lipids, and a corresponding increase in the amount of linoleic acid. (Browse et al., 1986). Thus, this mutant had the properties expected of a mutant deficient in a linoleic acid desaturase from the prokaryotic pathway which is primarily responsible for the synthesis of lipids in green tissues.

An unusual property of the fadD mutants was that they were very deficient in linoleic acid content when grown at temperatures above about 22 °C but had almost normal fatty acid composition when grown at temperatures below about 18 °C (McCourt et al., 1987). Since it was very unlikely that several independently isolated mutations would all give rise to a temperature conditional phenotype, it was concluded that a second desaturase must be partially responsible for desaturating linoleic acid to linolenic acid in green tissues. Therefore, the fadD mutant was remutagenized with ethylmethane sulfonate, self-fertilized to produce a segregating population of mutagenized plants (designated the M2 generation), and this population was screened for a mutant which was deficient in linolenic acid in green tissues at low temperatures. A mutant

with this property was isolated and the mutation responsible for this effect was designated the fadE locus (Somerville and Browse, unpublished).

Isolation of the Linoleic Acid Desaturase Gene from Canola

The following example was used to isolate the structural coding sequence from the fad3 region. The method described herein could equally have been used to isolate either the fadD or fadE region.

In order to approximately locate the fad3 mutation of the genetic map of *Arabidopsis*, a sexual cross was made between the fad3 mutant line BL1 and the multiply marked mutant line W1 (Hugly et al., 1991). The F1 hybrids from this cross were permitted to self-fertilize and the resulting F2 plants were scored for both the segregating genetic markers and the altered fatty acid composition. The results of this analysis indicated that the fad3 mutation was located on chromosome 2 near the marker erecta. In order to obtain a more accurate map position by RFLP mapping, a second sexual cross was made between the fad3 mutant line BL1 and the Niederzenz race of *Arabidopsis*. The F1 progeny were permitted to self-fertilize to produce the F2 generation. 137 F2 plants were grown during 3 weeks at 22° C (100 $\mu E/m^2/s$) in order to produce fully expanded rosettes, and a few leaves (representing a total weight of 0.2-0.5 g per plant) were harvested from each plant in order to prepare DNA from them.

The leaves were frozen in liquid nitrogen, and ground in dry ice, using a mortar and a pestle. For each sample, the frozen powder was transferred to a microfuge tube and an equal amount of 2 X CTAB buffer (2% cetyltrimethyl ammonium bromide (CTAB), 100 mM Tris-HCl pH 8, 20 mM EDTA, 1.4 M NaCl, 1% polyvinylpolypyrrolidone (PVP) 40,000) was added. The tubes were left at room temperature for 5 min to allow the powder to thaw. The homogenate was extracted once with a mixture of chloroform-isoamyl alcohol (24:1, v/v), and 1/10 vol of 10 X CTAB (10 % CTAB, 0.7 M NaCl) buffer was added to the aqueous phase, which was then

reextracted with an equal volume of chloroform isoamyl alcohol (24:1, v/v). The aqueous phase was transferred to a fresh microfuge tube and 1.5 vol of CTAB precipitation buffer (1% CTAB, 50 mM Tris-HCl pH 8, 10 mM EDTA) was added. The DNA was allowed to precipitate for 12 hr at 4 degrees, and collected by centrifugation (5 min at 10 000g). The DNA was resuspended in 100 µl of 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 M NaCl, and 100 µg/ml RNase A and incubated at 50°C for 30 min. The DNA was precipitated by adding 2.2 vol of ethanol and incubating on ice for 20 min. The DNA was collected by centrifugation and the pellet was washed once with 1 ml of 70% ethanol, dried under vacuum for 3 min and resuspended in 10 µl of distilled water. The DNA was stored at -20°C until use.

The 137 plants were grown to maturity and their seeds were collected individually. The fatty acid composition of 10 individual seeds from each of the F2 plants was measured as described by Browse et al (1986) in order to score the fad3 phenotype of each plant. Each seed was incubated in 1 ml of 1N HCl in methanol for 1h at 80°C. The tubes were cooled to room temperature and 1 ml of 0.9 % NaCl plus 0.3 ml of hexane were added. The tubes were agitated by vortexing and the phases separated by centrifugation (300xg for 5 min). The hexane phase was saved, evaporated under a stream of nitrogen, and the fatty acid methyl esters were dissolved in 50 µl hexane. An aliquot (2 µl) was injected onto the gas chromatograph and the fatty acid methyl esters separated and quantitated by flame ionization as described (Browse et al., 1986).

The DNA samples (1 µg) were then cut with the appropriate restriction enzyme (EcoR1 for the marker # 220, Bgl2 for the marker ASA2) using a concentration of 1XKGB buffer (Sambrook et al, 1989), 5 units of the restriction endonuclease and 100 µg/ml BSA. The volume of each sample was 10 µl and the incubation was performed at 37 °C for 4 h. The fragments were resolved by agarose gel electrophoresis (0.8 % agarose

in 1X TAE buffer; Sambrock et al., 1989) and transferred to nylon filters (hybond N+), using the alkaline transfer method as described by the manufacturer. The nylon filters were probed (according to Church and Gilbert, 1984) with radioactively labelled fragments of DNA (Sambrock et al., 1989) corresponding to known RFLP markers which had previously been mapped in the approximate vicinity of the fad3 locus on chromosome 2. The RFLP markers 220 (Chang et al 1988) and ASA2 were found to map close to the fad3 locus. Analysis of the pattern of recombinants (Table 1) indicated that both ASA2 and 220 were located on the same side of the fad3 locus at distances of 0.4 and 2.2 centimorgans (cM), respectively.

Table 1

# of plants	220	ASA2	fad3
67	H	H	+/-
30	L	L	-/-
34	N	N	+/+
3	H	N	+/+
1	L	H	+/-
1	N	H	+/-
1	H	H	-/-

Table 1 shows the genotype of the F2 plants used for mapping the fad 3 locus. L is for Landsberg (background of the fad 3 mutant), N is for Niederzenz, H for heterozygous. A total of 137 F2 plants were analyzed. The number of recombinant plants between fad3 and 220 or ASA2 was 6 and 1 respectively.

In order to isolate the region of the chromosome containing the fad3 locus, the RFLP markers 220 and ASA2 were used as hybridization probes to screen several yeast artificial chromosome (YAC) libraries. (Grill

and Somerville, 1991; Ward and Jen, 1990). The YAC filters were prepared according to Grill and Somerville (1991). The library was replicated onto nylon filters disposed on petri dishes of SC – – (synthetic complete medium minus tryptophan and uracil; Sherman et al., 1986). The cells were allowed to grow for 12 h at 30°C, and the filters were transferred for 15 min on a Whatman 3MM paper saturated with 1 M sorbitol, 50 mM DTT, 50 mM EDTA (pH 8).

The cell wall of the cells was then digested with lyticase, by incubating the filters on a Whatman paper saturated with 1M sorbitol, 50 mM EDTA and 2 mg/ml lyticase (Sigma Co., St. Louis,MO) for 12 h at 30°C. The filters were then transferred on a Whatman 3MM paper saturated with 0.5 M NaOH, 1.5 M NaCl for 15 min, neutralized with 0.5 M Tris-HCl pH 8 for 15 min and quickly rinsed in 2XSSC (SSC is 10mM sodium citrate, 150mM NaCl, pH 7). The filters were allowed to dry, and were transferred to a vacuum oven at 80°C for 1 h. They were subsequently hybridized according to Church and Gilbert (1984), with probes labelled with ^{32}P according to Sambrook et al. (1989).

The DNA of RFLP probe 220 was prepared from 100 ml of liquid culture lysate using the lambdasorb procedure (Promega Corp., Madison, WI); the cDNA encoding ASA2 was excised from the original plasmid (pKN140C; obtained from Dr. G. Fink, Whitehead Institute, Cambridge, MA) with Hind3 and cloned into the Hind3 site of pBLUESCRIPT. The plasmid DNA was then purified by Cesium chloride gradients according to Sambrook et al (1989), digested with Hind3 and the DNA insert was gel purified twice by electroelution according to Sambrook et al (1989).

In order to probe the libraries, the whole DNA from RFLP220 was used as a hybridization probe. By contrast, only the DNA insert of ASA2 was used as a probe. The RFLP probe 220 hybridized to YAC

EG4E8 and EG9D12. The probe ASA2 hybridized to YACs EW15G1, EW15B4 and EW7D11.

In order to determine if these YACs contained all of the DNA between RFLP220 and ASA2, small regions of DNA from the ends of the inserts in EG4E8 and EW15G1 were prepared by inverse PCR (Grill and Somerville, 1991). For that purpose, DNA was prepared from the appropriate YAC clones. The clones (single colonies) were grown to saturation in SC– – liquid cultures, and 1 ml of these cultures was used to inoculate 40 ml liquid cultures (in SC– – medium) that were allowed to grow for 16 h at 30°C. The cells were collected by centrifugation, washed once in 1 M sorbitol, 50 mM EDTA, resuspended in 200 μl of 1 M sorbitol, 50 mM EDTA, 100 mM sodium citrate pH 5.8, 2 mM β-mercaptoethanol and 2 mg/ml lyticase, and incubated 2 h at 30 °C.

Next, 350 μl of 2XCTAB buffer was added and the DNA was purified as described above. DNA (5 μg) of each clone was digested separately with HincII, AluI, EcoRV and RsaI (in 1XKGB buffer, at 37 °C for 4 h; final volume: 50 μl). The reactions were stopped by heating at 65 °C for 15 min, extracted once with one volume of phenol saturated with TE pH 8, followed by an extraction with 1 volume of chloroform - isoamyl alcohol mixture (24:1, vol/vol). The DNA was recovered by ethanol precipitation and resuspended in sterile distilled water. The ligation reactions were performed using 300 ng of DNA in a final volume of 50 μl. The reactions were carried out in 50 mM Tris-HCl pH 7.4, 10 mM MgCl2, 1 mM DTT,1.2 mM ATP with 1 U of ligase, for 2 h at 20 °C, and stopped by heating at 68 °C for 30 min.

The PCR reactions were carried out as follows: The buffers used were the ones indicated by the suppliers except for the Perkin Elmer enzyme for which the reaction was supplemented with an additional 1.4 mM $MgCl_2$ (final concentration 2.9 mM Mg). The dNTP final concentration

was 125 µM when the Perkin Elmer enzyme was used and 200 µM with the Taq polymerases from other sources. In all cases, 100 ng of each oligonucleotide was used. The final volume was 100 µl. When no product was obtained, the reactions were carried out again in the same conditions except that formamide was added to a final concentration of 3 %.

The left end was amplified from the ligation products of the EcoRV and RsaI digests, using the oligonucleotides EG1 (GGCGATGCTGTCGGAATGGACGATA) (SEQ. ID NO. 3) and EG2 (CTTGGAGCCACTATCGACTACGCGATC) (SEQ. ID NO. 4).

The right end of the clones obtained from the EG library was amplified from the ligation products of the AluI and HincII digests, using the oligonucleotides EG3 (CCGATCTCAAGATTACGGAAT) (SEQ. ID NO. 5) and EG4 (TTCCTAATGCAGGAGTCGCATAAG) (SEQ. ID NO. 6).

The right end of the clones obtained from the EW YAC library was amplified using the oligonucleotides H1 (AGGAGTCGCATAAGGGAG) (SEQ. ID NO. 7) and H2 (GGGAAGTGAATGGAGAC) (SEQ. ID NO. 8), using the same cycle conditions as above, except that the annealing temperature was reduced to 50 °C.

After the reactions were completed, 5µl of each mixture were electrophoresed on an agarose gel to separate the amplification product from primers. The slice of agarose that contained the amplified band was excised from the gel and melted in 1 ml of distilled water. Large amounts of product could then be produced, by reamplification of 5 µl of the melted slice. The PCR products were then purified by electroelution or by using GeneClean (Bio101) and used as hybridization probes to probe filters containing the isolated YAC DNA restricted by several enzymes. The probe made from the right end of EW15G1 hybridized to EG4E8 and similarly, a probe from the right end of EG4E8 hybridized to EW15G1.

Thus, it was concluded that the YACs EG4E8 and EW15G1 contained all of the DNA in the region of the chromosome between RFLP220 and ASA2.

The size of the YAC clones was estimated by field inversion electrophoresis (CHEF, Vollrath and Davis, 1987). High molecular weight DNA was prepared as follows: the yeast cells which contained the YAC clones were grown and treated with lyticase as for preparing DNA as described above. The spheroplasts were then resuspended in an equal volume of 1M sorbitol, 50 mM EDTA, 1 % low melt agarose at 37°C. The mixture was poured in a mould (Biorad) which was set on ice to allow the agarose to harden.

The resulting plugs were incubated for 12 h in 0.5 M EDTA pH 9, 1% lauryl sarcosine 1 mg/ml Proteinase K at 50°C. The plugs were subsequently washed twice in 50 mM EDTA and stored at 4°C until use. The CHEF gel was run in 1XTBE for 16 h at 200 V, with a switching interval of 20 s; the temperature of the buffer was maintained at 14 °C during the run. The sizes of the YACs were determined by comparison with a lambda ladder and the yeast chromosomes, and were as follows: EG4E8, 90 kb; EG9D12, 190 kb; EW15G1, 90 kb; EW15B4, 70 kb, EW7D11, 125 kb. These sizes permitted us to roughly determine a correspondence between physical and genetic distances: the distance that separates 220 from ASA2 cannot exceed 180 kb, the sum of the size of the 2 YACs EG4E8 and EW15G1. Since the corresponding genetic distance is 1.7 cM, one can roughly estimate that, in this particular cross and in this particular region of the genome, the value of 1 cM is close to 100kb. Thus, since the fad3 gene maps only 0.4 cM away from ASA2, the corresponding physical distance should be close to 40 kb. We then concluded that fad3 was probably located on the YAC EW7D11, which is the largest YAC hybridizing with ASA2. See Figure 1.

In order to test the possibility that the YAC EW7D11 carried the fad3 gene, the YAC was used to probe a cDNA library made from developing seeds of Canola (*Brassica napus* L.). Even though the YAC was isolated from Arabidopsis, the fact that Arabidopsis and *B. napus* are both members of the family Cruciferae led us to predict that the homologous genes from these two species would be sufficiently identical at the nucleotide sequence level so that the Arabidopsis gene would hybridize to the *B. napus* gene. We also assumed that, because it catalyzes a chemically similar reaction to the stearoyl-ACP desaturase, it would be expressed at similar moderately high levels in developing seeds (Shanklin and Somerville, 1991). Since EW7D11 contained only about 0.2% of the total genome, we expected it to contain only about 2 moderately abundantly expressed genes (i.e., genes in which the mRNA is between 0.1 and 0.01% of total mRNA).

DNA of YAC EW7D11 was isolated as follows: high molecular weight DNA was prepared from the yeast cells that contained the YAC EW7D11 as described above, and several preparative low-melt agarose CHEF gels were run in 1XTBC buffer (same as TBE except that CDTA was substituted for EDTA). The slices that contained the YAC were excised from the gels and pooled. Three slices were melted at 65°C and extracted with an equal volume of phenol saturated with TE. The aqueous phase was saved and reduced to 0.5 ml by repeated extractions with isobutyl alcohol. The remaining agarose was removed by several phenol extractions, followed by two chloroform-isoamyl alcohol extractions. The DNA was precipitated by adding 2 µg of linear acrylamide as a carrier plus 10 µl of 5M NaCl and 1.1 ml of ethanol, and incubating 20 min at 0 °C. The DNA pellet was recovered by centrifugation, washed in 70% ethanol, dried under vacuum and resuspended in 50 µl of distilled water. The DNA (50 ng) was radioactively labelled and used to probe a cDNA library in λgt11.

The nitrocellulose filters were processed as described in Sambrook et al (1989). Duplicate filters were used, and the films were exposed 5-7 days in order to obtain a good signal. From among 200,000 plaques screened in this way, 31 hybridized to EW7D11. Among these 31 clones, 17 were homologous to each other, as checked by cross hybridization in stringent conditions. The size of the inserts in the 17 clones was estimated and the clone with the largest cDNA was retained for further analysis. A small scale preparation of this phage was prepared using the lambdasorb method, and the insert was excised by restricting with EcoR1. This insert was ligated into a pBLUESCRIPT II vector linearized with EcoRI, and the ligation mixture was used to transform E. coli strain DH5α.

One of the recombinant clones was designated pBNDES3 (Figure 2), and retained for sequencing. The sequence was determined on both strands, using the sequenase enzyme, (US Biochemicals, Cleveland, OH) according to the instructions provided by the supplier. The nucleotide sequence of the insert in pBNDES3 is presented as Figure 3. The deduced amino acid sequence of the largest open reading frame in the nucleotide sequence is also shown in Figure 3.

Comparison of the deduced amino acid sequence of the 383 amino acid open reading frame in clone pBNDES3 against the known sequences in GenBank release 70 was performed using the FASTA program (Lipman and Pearson, 1985). This analysis revealed that the sequence from pBNDES3 had a region of significant homology to a previously characterized desaturase gene from the cyanobacterium Synechocystis (Figure 4). (Wada et al. 1990). This was considered suggestive evidence that the clone pBNDES3 encoded a desaturase which was probably the fad3 structural coding sequence product. This was subsequently confirmed by a genetic complementation experiment.

The cDNA was cloned into plant transformation vector pBI121 (Figure 5) under the control of the CaMV35S promoter to construct pTiDES3 (Figure 6). Plasmid pTiDES3 was introduced into an *Agrobacterium tumefaciens* strain which also carried an Ri plasmid and this was used to produce transgenic rooty tumors from both wild type Arabidopsis and the fad3 mutant. Transgenic tissue was selected for antibiotic resistance to confirm the presence of the pTiDES3. Fatty acid methyl esters were then prepared and examined by gas chromatography to determine the profile of fatty acids being produced in the tissue. The levels of linolenic acid increased, demonstrating that the cDNA on pTiDES3 can complement the fad3 mutation. These results, which are described in detail in Example 1 below, confirm the identity of the cDNA as encoding a linoleic acid desaturase.

The isolation of a plant structural coding sequence provides those skilled in the art with a tool for the manipulation of gene expression by the mechanism of antisense RNA. The technique of antisense RNA is based upon introduction of a chimeric gene which will produce an RNA transcript that is complementary to a target gene (reviewed in Bird and Ray, 1991). The resulting phenotype is a reduction in the gene product from the endogenous gene. The portion of the gene which is sufficient for achieving the antisense effect is variable in that numerous fragments or combinations thereof are likely to be effective. Various portions of the structural coding sequence of linoleic acid desaturase isolated either from cDNA or genomic clones are likely capable of reducing linolenic acid levels in plants by reduction in levels of linoleic acid desaturase levels. An example of using an antisense oriented linoleic acid desaturase structural coding sequence is set out in Example 2.

Polyadenylation Signal

The 3' non-translated region of the double stranded DNA molecule of the present invention contains a region that functions in plant cells to promote polyadenylation to the 3' end of the RNA sequence. Any such regions can be used within the scope of the present invention. Examples of suitable 3' regions are (1) the 3' transcribed, non-translated regions containing the polyadenylated signal of *Agrobacterium* tumor-inducing (Ti) plasmid genes, such as the nopaline synthase (NOS) gene, and (2) 3' regions of plant genes like the soybean storage protein genes and the small subunit of the ribulose-1,5-bisphosphate carboxylase (ssRUBISCO) gene. An example of a preferred 3' region is that from the NOS gene, described in greater detail in the examples below.

Plant Transformation/Regeneration

Any plant which can be transformed to contain the double-stranded DNA molecule of the present invention are included within the scope of this invention. Preferred plants which can be made to have increased or decreased linolenic acid content by practice of the present invention include, but are not limited to sunflower, safflower, cotton, corn, wheat, rice, peanut, canola/oilseed rape, barley, sorghum, soybean, flax, tomato, almond, cashew and walnut.

A double-stranded DNA molecule of the present invention containing the functional plant linoleic acid desaturase gene can be inserted into the genome of a plant by any suitable method. Suitable plant transformation vectors include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed, e.g., by Herrera-Estrella (1983), Bevan (1984), Klee (1985) and EPO publication 120,516 (Schilperoort et al.). In addition to plant transformation vectors derived from the Ti or root-inducing (Ri) plasmids of *Agrobacterium*, alternative methods can be used to insert the DNA constructs of this invention into

plant cells. Such methods can involve, for example, the use of liposomes, electroporation, chemicals that increase free DNA uptake, free DNA delivery via microprojectile bombardment, and transformation using bacteria, viruses or pollen.

A plasmid expression vector, suitable for the expression of the linoleic acid desaturase gene in monocots is composed of the following: a promoter that is specific or enhanced for expression in the lipid storage tissues and a 3' polyadenylation sequence such as the nopaline synthase 3' sequence (NOS 3'; Fraley et al., 1983). This expression cassette may be assembled on high copy replicons suitable for the production of large quantities of DNA.

A particularly useful *Agrobacterium*-based plant transformation vector for use in transformation of dicotyledonous plants is plasmid vector pMON530 (Rogers, S.G., 1987). Plasmid pMON530 (see Figure 7) is a derivative of pMON505 prepared by transferring the 2.3 kb StuI-HindIII fragment of pMON316 (Rogers, S.G., 1987) into pMON526. Plasmid pMON526 is a simple derivative of pMON505 in which the SmaI site is removed by digestion with XmaI, treatment with Klenow polymerase and ligation. Plasmid pMON530 retains all the properties of pMON505 and the CaMV35S-NOS expression cassette and now contains a unique cleavage site for SmaI between the promoter and polyadenylation signal.

Vector pMON505 is a derivative of pMON200 (Rogers, S.G., 1987) in which the Ti plasmid homology region, LIH, has been replaced with a 3.8 kb HindIII to SmaI segment of the mini RK2 plasmid, pTJS75 (Schmidhauser & Helinski, 1985). This segment contains the RK2 origin of replication, oriV, and the origin of transfer, oriT, for conjugation into *Agrobacterium* using the tri-parental mating procedure (Horsch & Klee, 1986). Plasmid pMON505 retains all the important features of pMON200 including the synthetic multi-linker for insertion of desired DNA fragments,

the chimeric NOS/NPTII'/NOS gene for kanamycin resistance in plant cells, the spectinomycin/streptomycin resistance determinant for selection in *E. coli* and *A. tumefaciens*, an intact nopaline synthase gene for facile scoring of transformants and inheritance in progeny and a pBR322 origin of replication for ease in making large amounts of the vector in E. coli. Plasmid pMON505 contains a single T-DNA border derived from the right end of the pTiT37 nopaline-type T-DNA. Southern analyses have shown that plasmid pMON505 and any DNA that it carries are integrated into the plant genome, that is, the entire plasmid is the T-DNA that is inserted into the plant genome. One end of the integrated DNA is located between the right border sequence and the nopaline synthase gene and the other end is between the border sequence and the pBR322 sequences.

When adequate numbers of cells (or protoplasts) containing the linoleic acid desaturase gene are obtained, the cells (or protoplasts) are regenerated into whole plants. Choice of methodology for the regeneration step is not critical, with suitable protocols being available for hosts from *Leguminosae* (alfalfa, soybean, clover, etc.), *Umbelliferae* (carrot, celery, parsnip), *Cruciferae* (cabbage, radish, rapeseed, etc.), *Cucurbitaceae* (melons and cucumber), *Gramineae* (wheat, rice, corn, etc.), *Solanaceae* (potato, tobacco, tomato, peppers) and various floral crops. See, e.g., Ammirato (1984); Shimamoto, 1989; Fromm, 1990; Vasil and Vasil, 1990.

<u>Uses of Linoleic Acid Desaturase</u>

The present invention can be used for any modification (either increase, decrease, or mere change) of the oil content of a plant or plant tissue. Linolenic acid is an important constituent of several membranes in plant cells.

One preferred method is to modify the oil content of the plant to improve the plant's temperature sensitivity. For instance, plants deficient in linolenic acid display reduced fitness at low temperature (Hugly and

Somerville, 1992). Also, increased linoleic acid content in vegetative tissues has been implicated as a factor in freezing tolerance in higher plants (Steponkus et al., 1990 and references therein). In a preferred embodiment, expression of the linoleic acid desaturase structural coding sequence can result in the genetic modification of higher plants to achieve tolerance to low environmental temperatures. Transformation with pTiDES3 demonstrates that linolenic acid levels can be increased by expression of this gene in a constitutive manner. Chilling or freezing injury in crops may be overcome by expression of this gene in vegetative or reproductive tissues by employing an appropriate promoter.

Linolenic acid, a polyunsaturated fatty acid, is also extensively used in the paint and varnish industry in view of its rapid oxidation. Flax seed is a predominant source of this oil. Higher quantities of this fatty acid in rapeseed or soybean will provide opportunities for using vegetable oils from these sources as a replacement for linseed (flax) oil. Expression of a linoleic acid desaturase structural coding sequence in seed tissue can result in a higher proportion of linolenic acid in the storage oil.

Linolenic acid is further a precursor in the biosynthesis of jasmonic acid, an important plant growth regulator. Linolenic acid is converted to jasmonic acid by introduction of an oxygen to the carbon chain by a lipoxygenase, followed by dehydration, reduction, and several β-oxidations (Vick and Zimmerman, 1984). The activity of jasmonic acid has been measured in terms of induction of pathogen defense responses. By application of free linolenic acid to plants, plant pathogen defenses can also be induced (Farmer and Ryan, 1992). A model has been proposed to explain the ability of free linolenic acid to exhibit the effects associated with jasmonic acid (Farmer and Ryan, 1992). It is hypothesized that all of the enzymatic activities which are required for the conversion of linolenic acid to jasmonic acid are constitutively present in the cell and the rate limiting

step in the production of jasmonic acid is the availability of free linolenic acid. A likely route for the production of the free linolenic acid is by the activity of a lipase in the plasma membrane.

It further has been observed that exogenous jasmonic acid can more powerfully activate defense responses than can wounding. This suggests that wounds cannot generate enough free linolenic acid to support high level production of jasmonic acid. The activity of the lipase or the availability of appropriate substrate for the lipase may be rate limiting upon wounding. By increasing levels of available substrate, increasing linolenic acid levels in the plasma membrane, it should be possible to enhance a plant's ability to respond to pathogens by allowing for a higher production of jasmonic acid. Expression of a linoleic acid desaturase structural coding sequence can result in a higher molar percent linolenic acid in the plasma membrane of a plant cell therefore enhancing the jasmonic acid signaling pathway. It is our intent to evaluate plants containing high linolenic acid levels in root and foliar tissues for their pathogen resistance.

It is also undesirable to have significant levels of linolenic acid in cooking oils. Linolenic acid is unstable during cooking and is rapidly oxidized. The oxidized products impart rancidity to the finished product. Rapeseed or soybean oil containing less than about 3%, and preferably 2% or less of linolenic acid is ideal for use as a cooking oil. By expression of the antisense of the structural coding sequence for linoleic acid desaturase, it is possible to reduce the linolenic acid content of these oils.

All higher plants have linolenic acid and, therefore, contain genes for linoleic acid desaturases. Because of the many examples in which genes isolated from one plant species have been used to isolate the homologous genes from other plant species, it is apparent to any one skilled in the art, that the results presented here do not only pertain to the use of the *B.*

napus fad3 gene, or to the use of the gene to modify fatty acid composition in *B. napus*. Obviously, the linoleic acid desaturases from many organisms could be used to increase linolenic acid biosynthesis and accumulation in plants and enzymes from any other higher plant or algae can serve as sources for linoleic acid desaturase genes. For example, since a YAC containing the *Arabidopsis* gene was used to isolate the *B. napus* gene, it is apparent that the insert in pBNDES3 could be used as a probe of genomic libraries for isolation of the corresponding full length genes from other plant species. It is also likely that the information contained in the sequence of this gene will be useful to clone other lipid desaturases genes.

Expression of a linoleic acid desaturase in a sense orientation may also allow for the isolation of plants with reduced levels of linolenic acid. This could be accomplished by the mechanism of co-suppression (Bird and Ray, 1991). The molecular mechanism of co-suppression is at this time poorly understood but occurs when plants are transformed with a gene that is identical or highly homologous to an allele found in the plants genome. There are several examples where expression of a chimeric gene in plants can result in a reduction of the gene product from both the chimeric gene and the endogenous gene(s). Those skilled in the art will recognize that the resulting decrease in linolenic acid would be a direct result of expression of the linoleic acid desaturase structural coding sequence and would be correlated to the linoleic acid desaturase activity in the transformed plant.

Linolenic acid levels in plant cells can also be modified by isolating genes encoding transcription factors which interact with the upstream regulatory elements of the plant linoleic acid desaturase gene(s). Enhanced expression of these transcription factors in plant cells can effect the expression of the linoleic acid desaturase gene. Under these conditions, the increased or decreased linolenic acid content would also be caused by a corresponding increase or decrease in the activity of the linoleic acid

desaturase enzyme although the mechanism is different. Methods for the isolation of transcription factors have been described (Katagiri, 1989).

The following examples are provided to better elucidate the practice of the present invention and should not be interpreted in any way to limit the scope of the present invention. Those skilled in the art will recognize that various modifications, truncations, etc. can be made to the methods and genes described herein while not departing from the spirit and scope of the present invention.

Example 1

Expression of fad3 gene to increase linolenic acid

To verify the assumption that the cDNA insert in pBNDES3 encodes a linoleic acid desaturase, both wild type and fad3 mutation *Arabidopsis* were transformed to contain the cDNA insert. In order to express the linoleic acid desaturase structural coding sequence (hereafter referred to as the "fad3 gene") in plant cells, the plasmid pBNDES3 was digested with XhoI and the ends were filled in with the Klenow fragment of DNA polymerase (Sambrook et al 1989). The cDNA insert was subsequently excised by digestion with Sac1 and ligated into the Sac1 and Sma1 sites of the binary Ti plasmid vector pBI121 (Clontech Laboratories), thereby replacing the GUS reading frame. The ligation reaction was carried out in 20 µl for 12 h at 16 °C using 100 ng of both insert and vector, and one unit of T4 DNA ligase. The ligation mixture was used to transform competent DH5α *E. coli* cells (prepared by the calcium chloride method, according to Sambrook et al, 1989), and transformants were selected on L-broth plates that contained 50 µg/µl Kanamycin. Alkaline minipreparations of recombinant clones were analyzed for the correct restriction pattern. One of these plasmids, designated pTiDES3, was used for further experiments.

This plasmid was electroporated (according to Mersereau and Pazour, 1990) into *Agrobacterium tumefaciens* strain R1000 which carries an Ri plasmid. The transformed bacteria were selected on kanamycin LB plates for 2 days at 30 °C. DNA minipreparations of several recombinant bacteria were performed and analyzed as described above to verify the presence of the construct.

Young flowering stems of wild type and the fad3 mutant of Arabidopsis were sterilized for 30 min in 10% commercial bleach, 0.02% Triton X100, and 2-cm explants that contained the flowering stem were infected with R1000 (pTiDES3) This was performed by dipping the sectioned extremity in a drop of an overnight culture of the appropriate *Agrobacterium* that was grown from a single colony in LB medium supplemented with 50 ug/ml Kanamycin.

The infected stems were cultured for two days on solid MSO medium (Gibco MS salts plus Gamborg B5 vitamins, 3% sucrose and 0.8% agar). At this time the stem segments were transferred for 5 weeks to MS0 medium containing 200 μg/ml cefotaxime to kill the bacterium. After approximately two weeks, most of the stem explants had developed rooty tumors resulting from transfer of parts of the Ri plasmid into cells of the stem explants. In order to identify the rooty tumors which had also received the binary Ti plasmid pTiDES3, approximately 24 rooty tumors from each treatment were transferred to MSO medium containing 50 μg/ml of kanamycin to select for the growth of those roots which had been cotransformed with the binary Ti plasmid; the medium contained also 200 μg/ml of cefotaxime to inhibit bacterial growth. Following a further period of growth for 2 weeks, fatty acid methyl esters were prepared (as described above) from the roots for analysis by gas chromatography. The results of these analyses are presented in Table 2.

Table 2. Genotype

mol% Fatty acid	wildtype pBI121	fad 3 pBI121	wildtype pTiDES3	fad3 pTiDES3
16:0	22.0±2.9	21.2±1.6	21.1±0.9	21.3±2.3
16:1	2.5±0.7	1.6±0.8	2.0±0.1	1.5±0.2
18:0	2.3±1.9	2.3±1.9	1.9±0.2	1.6±0.4
18:1	3.8±1.3	5.9±2.6	7.7±2.0	9.1±2.0
18:2	37.3±3.7	62.2±5.9	15.7±11.7	24.4±14.9
18:3	31.9±4.5	6.7±0.7	51.3±10.9	42.1±15.5

Table 2 shows the fatty acid composition of transgenic roots. The transgenic roots resulting from infection of wild type or the fad3 mutant with *A. tumefaciens* R1000 carrying the vector (pBI121) or the plasmid pTiDES3 were grown in the presence of kanamycin (50 g/ml) for three weeks to identify the roots which had been cotransformed with one of these plasmids. The fatty acid composition of the roots was determined as previously described (Browse et al., 1986). The abbreviations used in Table 2 are as follows: 16:0, palmitic acid; 16:1, palmitoleic acid; 18:0, stearic acid; 18:1, oleic acid; 18:2, linoleic acid; 18:3, linolenic acid. The values presented are the mean ± SD (n=12).

From these results it can be seen that the production of rooty tumors containing pBI121 on wild type *Arabidopsis* or the fad3 mutant had no effect on the fatty acid composition over non-pBI121 containing wild type *Arabidopsis* or fad3 mutant. By contrast, transformation of the fad3 mutant with the plasmid pTiDES3 resulted in large increases in the content of linolenic acid. In contrast to the linolenic acid content of 6.7 +/- 0.7% in the fad3 mutant transformed with pBI121, the presence of pTiDES3 resulted in accumulation of 42.1% of the fatty acids as linolenic acid. The increased content of linolenic acid was accompanied by a

decrease of corresponding magnitude in the content of linoleic acid. Thus, it is clear that the fad3 gene encodes a linoleic acid desaturase. Introduction of the fad3 gene into wild type tissues also resulted in significantly increased accumulation of linolenic acid and a corresponding decrease in linoleic acid (Table 2). Thus, it is apparent from these results that the linoleic acid content of plant tissues can be increased by high level expression of a linoleic acid desaturase. In the present embodiment, the fad3 gene was placed under transcriptional control of the constitutive high level CaMV 35S promoter carried on pBI121. The implication from these results is that expression from this promoter raised the level of expression of the fad3 gene to levels higher than are normally achieved by expression from the endogenous fad3 promoter. The results presented here indicate that the fad3 gene has significant utility in genetic modification of higher plants to elevate linolenic acid levels.

Example 2

Antisense expression of fad3 gene to decrease linolenic acid levels

In order to decrease the linoleic acid desaturase activity by genetic engineering methodology, the cDNA insert of pBNDES3 was cloned into plant expression cassettes in an antisense orientation. A 959bp BglII restriction fragment of pBNDES3 was used in the antisense expression vectors. The fragment is from 152 nucleotides downstream of the initiating methionine codon of the cDNA to a second BglII restriction site that is located near the C-terminus of the coding region. 189 nucleotides of the coding region are excluded from this fragment. Triple ligations were performed with the fad3 gene fragment to construct two separate plant expression cassettes.

A seed specific expression cassette was constructed by insertion of the BglII fragment of pBNDES3 in an antisense orientation behind the soybean promoter for the α' subunit of β-conglycinin (7S promoter). A

975bp HindIII to BglII fragment containing the 7S promoter derived from pMON529 was prepared by digesting with BglII for 30min at 37 °C followed by addition of Calf Intestinal Alkaline Phosphatase (CIAP) (Boehringer Mannheim). The reaction was allowed to proceed for 20min followed by purification of the linearized DNA using the GeneClean (Bio 101) purification system. The DNA was then digested with HindIII. A fragment derived from pMON999 containing the Nopaline synthase 3' region and the pUC vector backbone was prepared by digestion with BamHI and treatment with CIAP. The DNA was purified by the GeneClean procedure and digested with HindIII. The fragment of pBNDES3 was prepared by digestion with BglII. The three fragments were purified by agarose gel electrophoresis and the GeneClean procedure. 50 to 200ng of the purified fragments were ligated for one hour at room temperature followed by transformation into the *E. coli* strain JM101. Resulting transformant colonies were used for plasmid preparation and restriction digestion analysis. Double digestion with BglII and NcoI was used to screen for transformants containing the fad3 gene in an antisense orientation. One clone was designated as correct and named pMON13801.

A second expression cassette was constructed to allow for constitutive expression of the antisense message in plants. A fragment containing the enhanced 35S promoter was prepared from pMON999 by restriction digestion with HindIII and BglII followed by treatment with CIAP as above. The correct sized fragment was obtained by agarose gel electrophoresis and the GeneClean procedure. The BglII to HindIII vector fragment and the BglII fragment of pBNDES3 which were purified above were used in this construction. Ligation, transformation and screening of clones were as described above. One clone was designated as correct and named pMON13802.

In both pMON13801 and pMON13802, the promoter, fad3 gene and the Nos 3' region can be isolated on a NotI restriction fragment. These fragments can then be inserted into a unique NotI site of the vector pMON17227 to construct glyphosate selectable plant transformation vectors. The vector DNA is prepared by digestion with NotI followed by treatment with CIAP. The fad3 containing fragments are prepared by digestion with NotI, agarose gel electrophoresis and purification with GeneClean. Ligations are performed with approximately 100ng of vector and 200ng of insert DNA for 1.5 hours at room temperature. Following transformation into the E. coli strain LE392, transformants were screen by restriction digestion to identify clones containing the fad3 expression cassettes. Clones in which transcription from the fad3 cassette is in the same direction as transcription from the selectable marker were designated as correct and named pMON13804 (FMV/CP4/E9, 7S/anti fad3/NOS) (Figure 8) and pMON13805 (FMV/CP4/E9, E35S/anti fad3/NOS) (Figure 9).

In preparation for transforming canola cells, pMON13804 and pMON13805 were mated into Agrobacterium ABI by a triparental mating with the helper plasmid pRK2013.

Seeds from the plants produced by transformation were analyzed for alterations in fatty acid profile. Fatty acid methyl esters (FAMES) were prepared from seed tissue and analyzed by capillary gas chromatography (Browse et al, 1986). For initial screening of plants, six seeds were pooled together from an individual plant. The seeds were crushed and FAMES extracts were made. Control plants, plants transformed with the selectable marker only (pMON17227), were also analyzed using the identical procedure. From the initial screen on pooled seed samples, several lines were identified which displayed a decreased level of linolenic acid. Lines with decreased levels of linolenic acid were

reanalyzed by determining fatty acid profiles from individual seeds. Four to twenty individual seed were analyzed from candidate lines and from selected control plants. The results of the FAMES analysis is summarized in Figure 9.

Figure 9 shows the levels of fatty acids expressed in molar percent of twenty individual seed of the transgenic line 13804-51 as compared to control seed. Panel A discloses oleic acid, panel B discloses linoleic acid and panel C discloses linolenic acid.

The data in Figure 9 demonstrate that antisense expression of a linoleic acid desaturase has significantly altered the fatty acid profile of the resulting seed tissue. The percent of linolenic acid has been reduced to a little over 2% of the total fatty acid in the seed tissue. The percent of linoleic acid has been reduced slightly and surprisingly, the percent of oleic acid in the seed has been increased to approximately 70%. This demonstrates the applicability of utilizing the fad3 gene to manipulate the fatty acid profile of crop plants.

In order to demonstrate that the alteration in the fatty acid profile of the FAMES extracted from total seed tissue would be reflected in the seed oil fraction, triglycerides from seeds of fad3 antisense plants were characterized. Total lipid extracts were made by pooling ten seeds and grinding in 2ml of methanol:chloroform:water (4:2:1). The homogenate was allowed to stand for 20min and then debris was pelleted and discarded. To the supernatant 400µl of chloroform:methanol (2:1), 640µl of chloroform and 740µl of water was added and vortexed. Phases were separated by centrifugation and the chloroform phase was recovered and dried under nitrogen. Samples were resuspended in 100µl of chloroform and 10µl was applied to silica gel G thin layer chromatography plates for separation. Two identical plates were prepared with one being charred after development to allow for alignment and location of spots to be analyzed on

the other plate. Plates were developed three times in petroleum ether:diethyl ether:acetic acid (90:10:1). One plate was sprayed with 50% sulfuric acid and heated in an oven at 90°C to allow for detection of lipids. Triglyceride fractions were identified as comigrating on the plate with purchased lipid standards (Sigma Chemical Co, cat #178-13). The charred plate was aligned with the identical plate and the triglyceride fractions were scraped from the plate. The fatty acids were transesterified to produce FAMES extracts for GC analysis by the same procedure as above. The fatty acid profiles of the triglyceride fractions are shown in Table 3 and demonstrate that this fraction have decreased linolenic acid.

TABLE 3

Transgenic line	Mol% 18:1	18:2	18:3
17227-10	44	30	15.3
17227-493	65	17	6.9
13804-47	58	21	4.3
13804-50	67	20	2.8
13804-76	59	19	5.0
13804-117	62	21	4.0

Table 3 compares the fatty acid molar percentages of triglyceride fractions from control and transgenic lines. These above results provide clear evidence that the fad3 gene can be used to decrease the levels of linolenic acid in the storage oil of plants. The gene provides a tool for the manipulation of the fatty acid profile of seed storage oil to improve the products derived from the oil.

A surprising result of this Example 2 is the effect the antisense fad3 gene has on the oleic acid content. The precise mechanism by which

antisense expression of a gene exerts an effect on the activity of an endogenous gene is unclear but is obviously a function of the homology of the sense and antisense gene products. Based upon the above experimental result, it would not be unreasonable to predict that the portion of the fad3 gene antisense message used contained a certain degree of homology with the genes providing the activity of one or more oleate desaturases. Therefore, a further advantage of the above invention is that it is possible that expression of a linoleic acid desaturase antisense message may exert an effect on oleate desaturase activity.

The unexpected nature of the reduction in oleic acid desaturase activity from the antisense fad3 plants is most apparent when one compares the fatty acid profiles from the antisense plants and the fad3 mutant of *Arabidopsis*. The levels of linoleic acid in the fad3 mutant plants increased when linoleic acid desaturase activity was eliminated by mutation. This indicates that the activity of the oleate desaturase was not highly effected by the loss of linoleic acid desaturase activity or by the accumulation of linoleic acid. In the fad3 mutant of Arabidopsis the level of linoleic acid increased when the level of linolenic acid decreased. However, a different pattern occurred in the antisense fad3 plants. In plants which exhibit a decreased percent of linolenic acid there is no corresponding increase, and is often a decrease, in the percent of linoleic acid. There is an increase in the percent of oleate in the antisense fad3 plants. This would indicate that oleate desaturase activity is depressed in these plants. The effects on the fatty acid profile by the fad3 mutation and the fad3 antisense expression are not equivalent, indicating that antisense expression of a linoleic acid desaturase can depress an oleate desaturase activity in plants.

Example 3

Modification of linolenic acid levels in soybean

The isolation of the fad3 gene from *B. napus* provides a tool to those with ordinary skill in the art to isolate the corresponding gene or cDNA from other plant species. There are many examples in which genes from one plant species have been used to isolate the homologous genes from another plant species. One such plant which could be improved upon by the modification of the level of linolenic acid is soybean.

Soybean oil typically contains linolenic acid at a level of 7-9% of the fatty acid in the oil. This level is undesirable because it promotes instability upon heating and imparts rancidity to the finished product. The levels of linolenic acid can be lowered by the expression of the soybean fad3 gene or cDNA in an antisense orientation in the developing seed. The following example describes one method for the isolation of a fad3 cDNA from soybean. However, similar procedures could be followed to isolate a genomic clone which could also be used to decrease the level of linoleic acid desaturase activity by antisense expression of a portion or all of the gene.

The fad3 gene from *B.napus* is used as a probe to screen a cDNA library constructed from soybean mRNA. In order to isolate a cDNA to be used in decreasing linolenic acid in seed, the optimal tissue to use for the isolation of mRNA is developing seed. There is, however, flexibility in the choice of methods and vectors which can be used in the construction and analysis of cDNA libraries (Sambrook et al, 1989). Procedures for the construction of cDNA libraries are available from manufacturers of cloning materials or from laboratory handbooks such as Sambrook et.al, 1989. Once a suitable cDNA library has been constructed from soybean, all or a portion of the fad3 cDNA from *B.napus* is labeled and used as a probe of the library. DNA fragments can be labeled for radioactive or non-radioactive screening procedures. The library is screened under suitable stringency.

Conditions are dependent upon the degree of homology between the fad3 gene of *B. napus* and soybean. Probe positive clones are plaque purified by standard procedures and characterized by restriction enzyme mapping and DNA sequence analysis. Clones are concluded to be soybean fad3 based upon data obtained from the sequence analysis or by expression in plants.

The entire clone or a portion thereof is placed down stream of a promoter sequence in an antisense orientation. Suitable promoters include seed specific promoters, such as the 7S (β-conglycinin) α'-subunit promoter, or less tissue specific promoters, such as the CaMV 35S promoter. An appropriate 3' non-translated region is placed downstream of the antisense cDNA to allow for transcription termination and for the addition of polyadenylated nucleotides to the 3'end of the RNA sequence. This expression cassette is then combined with a selectable or scorable marker gene and soybean cells are transformed by free DNA delivery (Christou et al, 1990) or an Agrobacterium based method of plant transformation (Hinchee et al, 1988). Plants recovered are allowed to set seed and mature seed are used for the production of FAMES by the procedures outlined above. The FAMES extracts are analyzed by gas chromatography to identify plant lines with reduced levels of linolenic acid in the seed.

Alternatives to the above methods may include but are not limited to the use of degenerate oligonucleotides as probes to screen the library. Degenerate oligonucleotide probes would be most optimally designed by choosing short segments of the fad3 amino acid sequence where the degeneracy of the genetic code is limited or by choosing sequences which appear to be highly conserved between the fad3 gene of *B. napus* and other known linoleic acid desaturases, such as the desaturase from the cyanobacterium *Synechocystis*. The oligonucleotides could be labeled and used to probe a soybean cDNA library. Alternatively, degenerate

oligonucleotides could be used as primers for the isolation of a portion or all of the soybean cDNA by PCR amplification.

Similar procedures could be used to isolate the homologous genes from other plant species. Another preferred plant species which could be improved upon by the modification of the level of linolenic acid is flax. Flax oil typically contains linolenic acid at a level of 45-65% of the fatty acid in the oil. This level is undesirable because it promotes instability upon heating and imparts rancidity to the finished product.

Example 4

Sense expression of fad3 to obtain reduced levels of linolenic acid

The cloning of the fad3 gene also provides a tool to decrease the levels of linolenic acid via the mechanism of co-suppression. The molecular mechanism of co-suppression occurs when plants are transformed with a gene that is identical or highly homologous to an allele found in the plants genome (Bird and Ray, 1991). There are several examples where expression of a chimeric gene in plants can result in a reduction of the gene product from both the chimeric gene and the endogenous gene(s). Therefore the fad3 gene product of *B. napus* may be reduced by transformation of *B. napus* with all or a portion of the fad3 cDNA which has been isolated. The resulting plant has reduced linoleic acid desaturase activity in tissues where the chimeric gene is expressed. The phenotype of reducing the linoleic acid desaturase activity is a reduction in the levels of linolenic acid. The mechanism of co-suppression could be applied to any plant species from which the fad3 gene is cloned and the plant species is transformed with fad3 in a sense orientation.

In order to reduce levels of linolenic acid by the mechanism of co-suppression, a plant transformation construct is assembled with the fad3 gene or cDNA in a sense orientation. The entire clone or a portion thereof is placed downstream of a promoter sequence in a sense orientation. Suitable

promoters include seed specific promoters, such as the 7S (β-conglycinin) α'-subunit promoter, or less tissue specific promoters, such as the CaMV 35S promoter. An appropriate 3' non-translated region is placed downstream of the fad3 gene to allow for transcription termination and for the addition of polyadenylated nucleotides to the 3' end of the RNA sequence. This expression cassette is then combined with a selectable marker gene and *B. napus* cells are transformed by an *Agrobacterium* based method of plant transformation. Plants recovered are allowed to set seed and mature seed are used for the production of FAMES which are analyzed by gas chromatography to identify plant lines with reduced levels of linolenic acid in the seed.

Example 5

Isolation of a chloroplast delta 15 desaturase from *Arabidopsis*

A fragment of 959bp was excised from the fad3 cDNA insert using the restriction endonuclease BglII, and labeled radioactively according to Feinberg and Vogelstein (1983). This fragment was used to probe a cDNA library from *Arabidopsis* thaliana as described above (Example 1) except that the hybridization temperature was 52° C. Several cDNA clones were positive, and one of them (pVA1) was further characterized. Its deduced amino acid sequence exhibited a strong homology with fad3 except at the N-terminus. The cDNA insert was placed under the control of the 35S promoter in the Ti vector pBI121, and the resulting construct, pBIVA12 was electroporated into Agrobacterium (C58 pGV3101). The bacterium was used to transform the Arabidopsis mutant fadD. For transformation, plants were grown at 22° C with a light intensity of 100/μE/cm^{-2}, until bolting (approximately 2 and 1/2 weeks). The stems (1mm-10mm long) were removed and the plants were inoculated with a drop of an overnight culture of the bacterium. The same operation was repeated 7 days afterwards.

The plants were then allowed to set seeds. The seeds were plated (2500 seeds per 150mm petri dish) on MSO plates that contained 50μg/ml kanamycin to select for plants that had integrated the construct. One transformant plant was obtained, and the fatty acids from its leaves were analyzed by gas chromatography (Table 4). The results obtained show that the pBIVA12 construct is able to reestablish the levels of linolenic and hexadecatrienoic acids in the fadD mutant at a level equal to or superior to the wild type. This demonstrates that pVA12 encodes the fadD gene.

TABLE 4

fatty acid	fadD	WT	FadD pBIVA12
16:0	13.0	14.0	14.9
16:1	4.9	4.3	4.2
16:2	8.7	0.5	0.3
16:3	3.0	13.2	9.5
18:1	3.3	2.3	1.2
18:2	36.4	10.9	5.8
18:3	30.8	54.6	63.7

Table 4 shows the complementation of the fadD mutant. Fatty acids were extracted from leaves of *Arabidopsis* according to Browse et al (1986) and were quantified (mol%) by gas chromatography. WT stands for the Columbia wild type.

Example 6

Isolation of a second chloroplast delta 15 desaturase from *Arabidopsis*

A fragment of 959 bp was excised from the cDNA insert using the restriction endonuclease BglII, and labelled radioactively according to Feinberg and Vogelstein (1983). This fragment was used to probe a cDNA library from *Arabidopsis*, exactly as described above (Example 5). Among the several positive clones obtained, the cDNA pVA34 was further characterized. Its deduced amino acid sequence exhibited 71.8% and 79.5% homology with fad3 and fadD, respectively. The N-terminus resembled a chloroplast transit peptide, meaning that this protein is likely to be localized to the chloroplast. The strong homology with fad3 and fadD suggests that the protein is also a delta 15 desaturase. Aside from fad3 and fadD, the only locus known to control delta 15 desaturation is the fadE locus, which controls a temperature-induced delta 15 desaturase. Therefore, it is likely that the cDNA contained within the clone pVA34 corresponds to the fadE locus.

Example 7

Linoleic desaturase homology to plant oleic desaturases

The linoleic desaturase genes are the first plant desaturases isolated whose proteins enzymatically perform the desaturation of an unsaturated fatty acid precursor. The reaction that linoleic desaturase performs and the cofactors it uses are likely to be very similar for the oleic desaturase reaction. Given the similar reactions, similar substrates and probably similar cofactors, it is likely that the oleic desaturase genes and proteins have homology to the linoleic desaturase genes and proteins. That the genes share homology is supported by the finding that antisense expression of the linoleic acid desaturase message results in higher oleic acids levels, which experimentally indicates homology between the linoleic and oleic desaturases. These factors indicate that the linoleic desaturase

protein and nucleic acid sequences provide useful information for isolating other lipid desaturase genes, particularly oleic desaturase genes.

a. Identification of unknown cDNA sequences in databases.

Random cDNA sequencing generates a large number of sequenced clones but provides no information about the function of the encoded proteins. Homology to known proteins is the quickest method for identifying the protein function encoded in the sequenced cDNA. However, homology searches are informative only when a homology with a previously characterized protein are found. A cDNA sequence that is not homologous to any known protein remains in the unknown function category. Thus the results functionally identifying the linoleic desaturases by sequence and by their ability to complement mutations in plant desaturase genes now provides a method for identifying the function and identity of random cDNA clones by their homology to the linoleic desaturases. Additionally oleic desaturases are identified by their homology with linoleic desaturases.

A TFASTA search of the GenBank and EMBL public data bases for genes encoding proteins homologous to the protein sequence of the linoleic desaturase fad3 has identified both linoleic desaturases and a second class of plant lipid desaturases likely to be oleic desaturases. In particular, sequences found in GenBank and EMBL and identified as T04093 and T12950 show significant homology to linoleic desaturases but show less homology than other linoleic desaturases. These sequences have 30% homology to fad3 and 56% similarity to fad3 linoleic desaturase (TABLE 5). The full length clone of these cDNAs is obtained by standard methods and is inserted into plant gene expression and transformation vectors and transformed into fad2 Arabidopsis mutants to confirm the identity of the oleic desaturase by genetic complemention as was described in the example with linoleic desaturase.

TABLE 5

Comparison of Fad3 and T04093 Protein Sequences

```
Percent Similarity: 52.381%   Percent Identity: 30.476%

                    .         .         .         .         .
fad3   101 GHGSFSDIPLLNSVVGHILHSFILVPYHGWRISHRTHHQNHGHVENDESW 150
                            |:|||:|||| :|:.||| || |  | :|.|| :
T04093   1 ................LIFHSFLLVPYFSWKYSHRRHHSNTGSLERDEVF 34

                    .         .         .         .         .
       151 VPLPEKLYKNLP.....HSTRMLRYTVPLPMLAYPIYLWYRSPGKEGSHF 195
           ||  ... .| ..       . .|::..||.: :|::|:|| :. .|:   .:
        35 VPKQKSAIKWYGKYLNNPLGRIMMLTVQF.VLGWPLYLAFNVSGR...PY 80

                    .         .         .         .         .
       196 NPYSSLFAPSERKLIATSTTCWSIMLATLVYLSFLVDPVTVLKVYGVPYI 245
           :.:.: | |... . . . . :. :: .:
        81 DGFACHFFPNAPIYNDRERSRYTSLMRVF*.................... 110
```

b. <u>Isolation of a oleic desaturase cDNA.</u>

The protein sequence of plant linoleic desaturases can be used to isolate oleic desaturases. The conserved regions between the linoleic desaturases and the DesA oleic desaturase are functionally important and are conserved in the plant oleic desaturase proteins as well. These conserved amino acid sequences provide a method of isolating plant oleic desaturases. There are several regions of the linoleic desaturase fad3 that are conserved in fadD, fadE and DesA. The consensus amino acid sequence is shown in Table 6, with the amino acids identical in all four proteins shown in capital letters. As described below, oligonucleotides designed to encode the amino acids sequences in the conserved regions are used to identify and isolate plant oleic desaturases.

TABLE 6

Fad3 Protein Sequence and Peptide Targets

```
MVVAMDQRSNVNGDSGARKEEGFDPSAQPPFKIGDIRAAIPKHCWVKSPLRSMSYVTRD
v.tplttp ...spseed..erfdpgapppf.laDIraaiPKhCwvKnpwksmsyVvrd
                                         DIraaiPKhCwvK
                                    (1a) DIraaiP
                                        (1b) aiPKhC
                                           (1c) KhCwvK

IFAVAALAMAAVYFDSWFLWPLYWVAQGTLFWAIFVLGHDCGHGSFSDIPLLNSVVGHIL
va.vfalaa.aayfnnW.lwPlyW.aqGTmfwalFVlGHDCGHgSFsndp.lNsvvGH.l
                WflwPlyWvaqGT    FVlGHDCGHgSF
           (2a) WflwPlyW    (3a) FVlGHD
           (2b) WflwP       (3b) VlGHDC
              (2c) wPlyW       (3c) GHDCGH
                  (2d) WvaqGT     (3d) CGHgSF

HSFILVPYHGWRISHRTHHQNHGHVENDESWVPLPEKLYKNLPHSTRMLRYTVPLPMLAY
hssilvPyHgWRisHrtHHqnhghvEnDesWhPl.ekiyknlpk.trmfrftlplpmlay
      PyHgWRisHrtHH      EnDesWvP
 (4a) PyHgW         (5a)  EnDesW
   (4b) HgWRisH        (5b) DesWvP
     (4c) WRisHrtHH
     (4d) WRisH
        (4e)  HrtHH

PIYLWYRSPGKEGSHFNPYSSLFAPSERKLIATSTTCWSIMLAT.LVYLSFLVDPVTVLK
pfylw.rspgk.gShyhpds.lF.pkerkdvltStacwtamaAl.lvcLnft.gpiqmlK

VYGVPYIIFVMWLDAVTYLHHHGHDEKLPWYRGKEWSYLRGGL.TTIDRDYG.IFNNIH
lygiPywifvmWldfvTylHHhghedklpwyrgkeWSylrggL.tTldrDYg.winnih
          WldavTylHH                 WSylrggL.tTidrDY
     (6a) WldavT                 (7a) WSylrggL
          (6b) TylHH                    (7b) L tTidrD
                                           (7c) TidrDY
```

```
HDIGTHVIHHLFPQIPHYHLVDATRAAKHVLGRYYREPKTSGAIPIHLVESLVASIK
HDIgtHviHHLfpqIPhYhLveAteaaKpvlGkyyrEpk.sgplplhLlesl.ksik
HDIgtHviHHLfpqIPhY
(8a) HDIgtH
  (8b) HviHHL
    (8c) HHLfpqI
     (8d) HLfpqIP
      (8e) LfpqIPhY

KDHYVSDTGDIVFYETDPDLYVYASDKSKIN*
.dhyvsdtGdvvyYeadp.lyg..s*
```

c. <u>Isolation of the fadC (fad6) Gene from *Arabidopsis thaliana*</u>

The fadC gene (also referred to as fad6) encodes a chloroplastic omega-6 desaturase.

The deduced amino acid sequences of the fad3 gene from *Brassica napus* and the fadD and fadE genes from *Arabidopsis thaliana* were compared with the DesA gene from *Synechocystis* (*Nature*, **347**:200, 1990). The sequence GHDCGH was determined to represent the most highly conserved region of these proteins. Consequently, a degenerate oligomer was designed that contains all the possible condons for the sequence GHDCGH. This oligomer has the following sequence: GGNCAYGAYTGYGGNCA.

An *Arabidopsis thaliana* cDNA phage library obtained from the laboratory of Dr. Ron Davis (*PNAS*, **88**: 1731-1735) was used to screen for desaturase genes. This library was made using material from all above ground plant parts.

Approximately 120,000 phage from the library were plated onto three plates and hybondN+ was then used to prepare three filters from each plate (*Molecular Cloning - A Laboratory Manual*, 2nd Edition. Eds. J. Sambrook, E. F. Fritsch, and T. Maniatis, Cold Spring Harbor Laboratory

Press, Cold Spring Harbor, New York 1989, hereafter "Sambrook"). Two filters from each plate were probed using the degenerate consensus oligomer which had been end-labelled with (32)P using T4 polynucleotide kinase. The hybridizations were performed in a solution that contained high amounts of tetramethylammonium chloride in order to minimize differences in the melting temperatures of the oligomers that together comprise the degenerate consensus oligomer. The hybridization solution had the following composition: 3 M tetramethylammonium chloride, 10 mM sodium phosphate pH 6.8, 1.25 mM EDTA, 0.5% SDS, 0.5% milk. Hybridization was carried out overnight at a temperature of 44ºC. Filters were then washed four times, 20 minutes each time, with 6 x SSC + 0.15% SDS at room temperature. Filters were then washed one time, for 30 minutes, with 4 x SSC + 0.1% SDS at room temperature. The filters were then exposed to film for two days.

The third set of filters that were made from each phage-containing plate were probed using DNA sequences from the three *Arabidopsis* desaturase genes that had already been identified: fad3, fadD and fadE. The fad3, fadD and fadE genes were labelled with (32)P and hybridized to the third set of phage filters in the following hybridization solution: 0.2 M NaCl, 20mM sodium phosphate pH 7.7, 2mM EDTA, 1% SDS, 0.5% milk, 10% dextran sulfate, 0.1% sodium pyrophosphate. Hybridization was carried out overnight at 65ºC. Filters were washed four times, 30 minutes per time, in 2 x SSC + 0.15% SD at room temperature and then for 45 minutes with 1 x SSC + 0.1% SDS at 65º C. The filters were then exposed to film for approximately two hours.

The two sets of filters that were probed with the degenerate consensus oligomer showed about 60 positive phage per plate (or about 180 total positive phage). Results from the third set of filters that were probed with the fad3, fadD and fadE genes indicated that only a small percentage

of the phage that hybridized to the consensus of oligomer contained the fad3, fadD or fadE genes.

Seventy-six of the phage that hybridized to the consensus oligomer, but not to the fad3, fadD or fadE genes, were plaque purified. The purified phage were then spotted onto bacteria growing on solid media on plates and allowed to form plaques. Several duplicate filters were then made of these plates (Sambrook). One of these filters was probed with the consensus oligomer, as described above. A second filter was probed with a mixture of the *Arabidopsis thaliana* fad3, fadD and fadE genes, as described above.

In order to determine which of the 76 phage contained the same cDNA inserts as which other phage, some of the filters were probed with cDNA inserts from some of the phage. In order to perform this experiment, the cDNA inserts from most of the phage were isolated by using oligomers that bound to DNA flanking the cDNA cloning site in the phage vector to isolate the cDNA sequences using the polymerase chain reaction (PCR). These cDNA sequences were labelled with (32)P (random hexamer labelling) and hybridized to the filters using the following hybridization solution: 30% formamide, 0.2M NaCl, 20mM sodium phosphate pH 7.7, 2mM EDTA, 1% SDS, 0.5% milk, 0.1% sodium pyrophosphate. The hybridizations were carried out for 14 hours at 65ºC. The filters were washed four times 15 minutes per wash, with 2 x SSC + 0.15% SDS at room temperature and were then exposed to film.

The combination of the high formamide concentration in the hybridization solution and the high hybridization temperature meant that only DNA sequences that were virtually identical would hybridize, allowing us to distinguish between nearly identical sequences. Several rounds of hybridizations using cDNA inserts from different phage were carried out until it had been determined which phage contained the same, or at least

extremely similar, cDNA inserts. On the basis of these experiments, we determined that all of the 76 phage contained one of four cDNA inserts. Sequence data was obtained from each of these four cDNAs. None of these cDNAs was found to be homologous to known desaturase genes, and so we feel that none of these four cDNAs is likely to encode a desaturase.

Since the number of phage that hybridized to the consensus oligomer was quite high (about 180 phage hybridized in the initial screen described above), we were not able to analyze all of the positive phage in the initial experiments. So, an attempt was made to identify phage that hybridized to the consensus oligomer but that did not contain the fad3, fadD of fadE genes or one of the four non-desaturase encoding clones that were identified in the first screen. In order to do this, between 500,000 and 1,000,000 phage from the library described above were plated onto 10 plates. Three filters were made from each plate (Sambrook). Two of these three sets of filters were then hybridized with (32) P labelled consensus oligomer as described above except that hybridization was carried out at 42ºC instead of at 44ºC. The third set of filters were hybridized with (32)P labelled DNA from the *Arabidopsis* fad3, fadD and fadE genes together with DNA from each of the four cDNA's identified in the first round of screening as hybridizing to the consensus oligomer but not encoding desaturases. This third set of filters were hybridized in: 30% formamide, 0.2 M NaCl, 20mM sodium phosphate pH 7.7, 2mM EDTA, 1% SDA, 0.5% milk, 0.1% sodium pyrophosphate at 65ºC. All three sets of filters were hybridized for 12 hours and then washed several times with 2 x SSC + 0.15% SDS at room temperature. The filters were then exposed to film.

Approximately 200 phage from each plate hybridized to the consensus oligomer. 50-60% of these phage also hybridized to fad3, fadD, fadE or to one of the four clones identified in the first screen. About 58 phage that hybridized to the consensus oligomer, but not to fad3, fadD,

fadE or one of the four previously identified clones, were plaque purified. The purified phage were then spotted onto a bacterial lawn growing on solid media on a petri plate and the phage were allowed to form plaques. Several filters were prepared from these plates and hybridized with (32)P labelled cDNA inserts from various of the newly purified phage, as described above. In this manner, all of the phage identified in this second round of screening were found to contain one of eight different cDNA inserts.

Sequence data was obtained from each of the eight cDNA's. One of the cDNA's, which was contained within only one of the phage, was found to have some sequence similarity of a known desaturase gene from cyanobacteria, the DesA gene. Further sequence information was obtained for this clone. This additional sequence showed very significant sequence similarity to the DesA gene, confirming that the clone contained a desaturase gene. The remainder of the cDNA contained within the clone was sequenced and compared with the sequences of other known desaturases. The new desaturase was 53.0% identical to DesA at the nucleotide level and 43.9%, 45.6% and 47.0% identical to *B. napus* fad3, *Arabidopsis* fadD and *Arabidopsis* fadE, respectively. As the gene contained within the clone was significantly more similar in sequence to the DesA gene (which is a delta-12 desaturase) than to fad3, fadD or fadE (which are omega-3 desaturases), the new desaturase was expected to be a delta-12 (= omega-6) desaturase.

The additional sequence data also indicated that this new desaturase gene contains a region that has only a one base pair mismatch to the desaturase consensus sequence described above. This mismatch means that the new desaturase has the sequence GHDCAH instead of GHDCGH.

A clone containing a full length cDNA for this gene was isolated and completely sequenced. This full length cDNA was sub-cloned

into the plant transformation vector pBII121 such that the gene is transcribed under the control of the 35S promoter. This construct was then used to complement the phenotype of a fadC mutant (*Plant Phys.* **90**: 522-529, 1989) of *Arabidopsis thaliana*, indicating that the gene encodes a chloroplastic omega-6 desaturase.

d. Proposed isolation of fad2

The most highly conserved peptide regions in the linoleic desaturases and the DesA desaturase were chosen as regions likely to be conserved in oleic desaturases. These 8 conserved regions are shown in TABLE 6. These regions were chosen on the following basis: These regions have areas highly conserved between the 3 linoleic desaturases and DesA, with at least 4 identical amino acids over a 10 amino acid span. Once a region was identified as conserved, the fad3 linoleic desaturase sequence was used as the amino acid sequence for the source of homology to identify oleic desaturases. This is because both fad3 and the non-plastid oleic desaturases are thought to be localized to the endoplasmic reticulum and are most likely to contain similar amino acid sequences.

Several peptide endpoints in each conserved area were chosen as the basis to subsequently design oligonucleotide probes for identifying the oleic desaturase gene. The peptide endpoints were chosen to be between 5 and 9 amino acids in length. The peptide end points were chosen to end on the conserved (identical) amino acids, and most often to begin on conserved amino acids. The rationale is that within the larger conserved area, some amino acid portions are more highly conserved than others, that 15 to 27 (5 to 9 amino acids) nucleotides is a good primer size for PCR, and that for PCR it is important that the 3' end of the primer matches the target, with the conserved (identical) amino acids the most likely to be present in the oleic desaturases. These 28 "oleic desaturase" peptide targets (Table 6) are the basis oligonucleotides that are designed for

hybridizing to the oleic desaturase cDNA sequences to identify and isolate the oleic desaturase cDNA clone.

Several possible methods for designing oligonucleotides and isolating the genes encoding the target peptide regions are known. For a discussion of designing degenerate oligonucleotides see *PCR Protocols - A Guide to Methods and Applications,* Eds M. A. Innis, D. H. Gelfand, J J Sninsky and T. J. White, Academic Press, San Diego, California, 1990; and Sanxxxxx The two most common screening methods using the oligonucleotides are screening cDNA libraries and PCR amplification of specific cDNAs. Gene probes from fad3, fadD and fadE are used under stringent hybridization conditions to identify these cDNAs and discard them in the screen for oleic desaturase cDNA clones. The method for using degenerate oligonucleotides to screen a cDNA library has been described in the example above demonstrating the isolation of the fadC oleic desaturase gene. An immature plant seed active in oil biosynthesis, generally 2 to 5 weeks after pollination, preferably about 3 to 4 weeks after pollination, of a plant such as Arabidopsis or canola is used as the source of mRNA for making cDNA. First strand cDNA is made from the isolated mRNA and hybridized under stringent conditions in solution to an excess of biotinylated fad3, fadD and fadE cloned cDNAs. The hybrids and biotinylated nucleic acids are removed with strepavidin and a second round of substraction is done to remove any remaining fad3, fadD and fadE sequences. The cDNA remaining in solution is used for PCR reactions. (For 5' RACE, see below, a polyA tail is added to the first strand cDNA 3' end).

A method that can readily evaluate a number of degenerate oligonucleotides probes is degenerate PCR (See chapters by Compton and by Lee and Caskey in *PCR Protocols*, cited above). In this method a degenerate set of oligonucleotides encompassing all the possible codon choices for the target peptide is synthesized (such degenerate

targets (Table 6) are the basis oligonucleotides that are designed for hybridizing to the oleic desaturase cDNA sequences to identify and isolate the oleic desaturase cDNA clone.

Several possible methods for designing oligonucleotides and isolating the genes encoding the target peptide regions are known. For a discussion of designing degenerate oligonucleotides see *PCR Protocols - A Guide to Methods and Applications,* Eds M. A. Innis, D. H. Gelfand, J J Sninsky and T. J. White, Academic Press, San Diego, California, 1990; and Sambrook. The two most common screening methods using the oligonucleotides are screening cDNA libraries and PCR amplification of specific cDNAs. Gene probes from fad3, fadD and fadE are used under stringent hybridization conditions to identify these cDNAs and discard them in the screen for oleic desaturase cDNA clones. The method for using degenerate oligonucleotides to screen a cDNA library has been described in the example above demonstrating the isolation of the fadC oleic desaturase gene. An immature plant seed active in oil biosynthesis, generally 1 to 5 weeks after pollination, preferably about 2 to 4 weeks after pollination, of a plant such as Arabidopsis or canola is used as the source of mRNA for making cDNA. First strand cDNA is made from the isolated mRNA and hybridized under stringent conditions in solution to an excess of biotinylated fad3, fadD and fadE cloned cDNAs. The hybrids and biotinylated nucleic acids are removed with strepavidin and a second round of substraction is done to remove any remaining fad3, fadD and fadE sequences. The cDNA remaining in solution is used for PCR reactions. (For 5' RACE, see below, a polyA tail is added to the first strand cDNA 3' end).

A method that can readily evaluate a number of degenerate oligonucleotides probes is degenerate PCR (See chapters by Compton and by Lee and Caskey in *PCR Protocols*, cited above). In this method a degenerate set of oligonucleotides encompassing all the possible codon

TABLE 7

Peptide Targets for Fad2 Cloning

Peptide sequence		Oligo sequence 5′ - 3′
1a	DIRAAIP	GAYATHMGNGCNGCNATHCC
1b	AIPKHC	GCNATHCCNAARCAYTG
1c	KHCWVK	AARCAYTGYTGGGTNAA
2a	WFLWPLYW	TGGTTYYTNTGGCCNYTNTAYTGG
2b	WFLWP	TGGTTYYTNTGGCCN
2c	WPLYW	TGGCCNYTNTAYTGG
2d	WVAQGT	TGGGTNGCNCARGGNAC
3a	FVLGHD	TTYGTNYTNGGNCAYGA
3b	VLGHDC	GTNYTNGGNCAYGAYTG
3c	GHDCGH	GGNCAYGAYTGYGGNCA
3d	CGHGSF	TGYGGNCAYGGNWSNTT
4a	PYHGW	CCNTAYCAYGGNTGG
4b	HGWRISH	CAYGGNTGGMGNATHWSNCA
4c-1	WRISHRTHH	TGGMGNATHTCNCAYMGNACNCAYCA*
4c-2		TGGMGNATHAGYCAYMGNACNCAYCA*
4d	WRISH	TGGMGNATHWSNCAY
4e	HRTHH	CAYMGNACNCAYCAY
5a	ENDESW	GARAAYGAYGARWSNTGG
5b	DESWVP	GAYGARWSNTGGGTNCC
6a	WLDAVT	NGTNACNGCRTCNARCCA
6b	TYLHH	RTGRTGNARRTANGT
7a-1	WSYLRGGL	ARNCCNCCNCKNARRTARCTCCA*
7a-2		ARNCCNCCNCKNARRTANGACCA*
7b	LTTIDRD	RTCNCKRTCDATNGTNGTNA
7c	TIDRDY	RTARTCNCKRTCDATNGT
8a	HDIGTH	RTGNGTNCCDATRTCRTG
8b	HVIHHL	NARRTGRTGDATNACRTG
8c	HHLFPQI	DATYTGNGGRAANARRTGRTG
8d	HLFPQIP	GGDATYTGNGGRAANARRTG
8e	LFPQIPHY	RTARTGNGGDATYTGNGGRAANA

* synthesize 4c and 7a in two pools each to limit the degeneracy

Oligos for 6a - 8e are the complement of the coding sequence

TABLE 8
Table of Oligomers for PCR RACE of fad2

Peptide #	Oligo Length	Fold Degeneracy	Similarity with L26296	Similarity in Last 10 n.t.
1a	20	384	75 %	80 %
1b	17	192	88	80
1c	17	32	65	80
2a	24	64	79	100
2b	15	48	73	80
2c	15	48	100	100
2d	17	128	76	90
3a	17	384	76	70
3b	17	384	82	80
3c	17	128	88	90
3d	17	384	82	70
4a	15	64	80	70
4b	20	192	75	90
4c	26	96*	81	80
4d	15	216	87	90
4e	15	192	87	80
5a	18	96	72	80
5b	17	96	76	80
6a	18	256	78	80
6b	15	192	93	100
7a	23	256*	78	60
7b	20	384	90	80
7c	18	192	94	90
8a	18	384	72	70
8b	18	192	89	80
8c	21	384	81	100
8d	20	192	80	90
8e	23	192	83	70

* done in two oligo pools

Table 7 shows the 28 peptide targets from the eight conserved regions and the 30 degenerate oligonucleotides derived from the peptide sequences. The degeneracy was kept to less than 516 fold, for those instances where more degeneracy occurred, by the use of deoxyinosine (Sambrook et al.) and by not including the last nucleotide in the last codon, and in two cases by the use of two subpools. Table 8 shows the amount of degeneracy for each designed oligonucleotide sequence and the amount of homology of the oligonucleotides to the *Arabidopsis* oleic desaturase fad2 (Accession No. L26296). Also shown in Table 8 is the percent homology in the last 10 nucleotides on the 3' end of each primer, since this region is most important for annealing and elongation under PCR conditions. It is expected that both 10 of 10 and 9 of 10 homology matches, and probably 8 of 10 homology matches in the 3' primer regions will serve as efficient PCR primers. Note that for oligonucleotide sets 1a through 5b (for 3' RACE) the strand direction is the same as the mRNA while for oligonucleotide sets 6a through 8e (for 5' RACE) the direction is opposite of the mRNA. Four oligonucleotides have a 10 of 10 match in the 3' position, 6 oligonucleotides match 9 of 10 in the 3' position and 12 match in 8 of 10 nucleotides in the 3' position. Oligonucleotides corresponding to peptides 2a, 2c, 2d, 3c, 4b, 4d, 6b, 7c, 8c, and 8d show 90% or greater homology in their last 10 nucleotides and anneal to the oleic desaturase gene and serve as primers to this gene. This demonstrates the validity of using the conserved regions of the plant linoleic desaturases and DesA to identify and isolate plant oleic desaturases.

The first round of PCR products are subjected to two rounds of subtraction using biotinylated fad3, fadD and fadE cloned cDNA to remove any hybridizing fad3, fadD and fadE sequences with strepavidin. This subtracted DNA is greatly enriched for fad2 sequences and depleted of fad3, fadD and fadE sequences. These 30 samples are run on agarose gels,

blotted and hybridized with pools of probe from the 30 samples. Pools of 5 of each of the 30 PCR samples are labeled with random primers and hybridized to the blots of the 30 samples, for a total of 6 blots hybridized with 6 pools of 5 probes. Additionally, a pool of fad3, fadD and fadE probe is hybridized to a duplicate blot. Bands that do not hybridize strongly to fad3, fadD and fadE but do cross hybridize to probe made from a different sample are strong candidates for fad2 as fad2 is likely to be the only DNA amplified in two or more independent PCR reactions. Positively hybridizing lanes identify samples to amplify by PCR using the same primers as in the initial reaction for 5 to10 cycles and the PCR products are cloned into plasmid vectors. The same probe that recognized the sample on the blot is used to screen the library and identify the hybridizing clone. Positive clones are sequenced and identified as fad2 clones by their homology but non-identity with fad3, and further characterized as described below.

In the event that fad2 sequences are not sufficiently enriched in one round of PCR to be identified, a second round of PCR is performed. If the lack of detection is due to insufficient amplification of fad2, then another round of PCR using the same primers on the subtracted PCR first round samples and the same simple screen as described above will identify fad2. If there are too many competing non-specific reactions then a second round of PCR using a different primer combination will remove non-specific amplifications and enrich for fad2. To further enrich for fad2 sequences each of the initial 30 PCR samples (one for each oligonucleotide in Table 7) after subtraction as described above, is subjected to a second round of PCR reactions using a different primer combination than the first reaction. One of the primers would be the same degenerate oligonucleotide primer as in the first PCR reaction. The second primer would now be from one of the 30 primers in Table 7 from the opposite class, ie, primers from 1a to 5b form matched sets with primers from 6a to 8e (primers 1a to 5b are in the sense

direction while primers 6a to 8e are in the antisense direction). For example, if oligonucleotide 1a was used initially, it is used again as one of the two primers and the second primer is each of the 6a to 8e oligonucleotides for a total of 11 separate PCR reactions. In total the 30 initial reactions result in 418 second cycle PCR reactions, a number easily handled by PCR technology. Essentially this second PCR cycle accomplishes a "nested" or sequential PCR reaction step after removing all the linoleic desaturases by the subtraction step. This increases the amplification as well as the specificity. Identification of samples containing fad2 are performed similarly as described above, with the 418 samples dot blotted onto 22 filters and probed with 21 pools of 20 samples and with a pool of fad3, fadD and fadE. Again, any sample that cross hybridizes with an independent probe sample and does not hybridize to fad3, fadD and fadE is a candidate for containing fad2 in the sample. If fad3, fadD and fadE hybridization is still present, another biotinylation/stepavidin subtraction should remove it. Positively hybridizing samples are run on gels, the band identified by hybridization and isolated for cloning. This second set of PCR reactions produces PCR products of a predictable size since both primers are within the coding region where little variation in size is expected. Thus the presence of a band of the expected size on a gel is diagnostic of fad2, particularly if hybridization of a blot of such a gel with a fad3, fadD and fadE probe indicates the band is not due to fad3, fadD and fadE contamination. After cloning the inserts in E. coli, the resulting plasmids containing the insert are identified by hybridization. They are sequenced and identified as oleic desaturases by their homology but non-identity with the linoleic desaturases, as in the examples described previously. The full length clone of these cDNAs is obtained by standard methods and inserted into plant gene expression and transformation vectors and transformed into *Arabidopsis* fad2 mutants to confirm the identity of the oleic

desaturase by genetic complemention as was described in the example with linoleic desaturase.

Thus in this approach to isolating the plant oleic desaturases, the total number of peptide regions is 8, comprised of 28 smaller peptide targets. This leads to set of 30 degenerate oligonucleotides, that are used in the PCR amplification and screening of the PCR products. Subtraction of interfering fad3, fadD and fadE sequences is used at several points. If necessary a second round of PCR reactions with paired internal primers gives extra amplification and specificity. This approach identifies the plant oleic desaturases, and the sequence of the isolated clones should confirm their identity by their homology to the plant linoleic desaturases as described. Thus a defined approach to isolating the plant oleic desaturases from the information about linoleic desaturases is presented here. The example given here is for Arabidopsis or canola oleic desaturases, but the approach is not limited to those plants as the oleic desaturases are probably highly conserved in most plants. Thus once one plant oleic desaturase is isolated, the sequence information is used to isolate the genes from other plant species by direct hybridization or by an approach similar to the one described here.

The following documents, which are cited in the above application, are hereby incorporated by reference.

Ammirato, P.V., et al. Handbook of Plant Cell Culture - Crop Species. Macmillan Publ. Co. (1984).

Bartlett, S.G., A.R. Grossman, & N.H. Chua. (1982) In Methods in Chloroplast Molecular Biology. Elsevier Biomedical Press, New York, pp 1081-1091.

Beachy, R. N., Chen, Z. L., Horsch, R. B., Rogers, S. G., Hoffmann, N. J., and Fraley, R. T. (1985) EMBO J, 4:3047-3053.

Benfey, P., Ren, L., and Chua, N.H. (1989) EMBO J, Vol.5, no.8, pp 2195-2202.

Bevan, M. (1984) Nucleic Acids Res. 12 (22): 8711-8721.

Bird, C. R., and Ray, J. A. (1991) Biotech. Gen. Engin. Rev. 9:207-227.

Bray, E. A., Naito, S., Pan, N. S., Anderson, E., Dube, P., and Beachy, R. N. (1987) Planta 172:364-370.

Browse J., McCourt P., Somerville C. R. (1985) Science 227:763-65.

Browse, J., P.J. McCourt, C.R. Somerville. (1986) Anal. Biochem. 152:141-146.

Browse, J., P. McCourt, C.R. Somerville. (1986b) Plant Physiol. 81:859-864 .

Browse, J. A., Slack, C. R. (1981.) FEBS Lett. 131:111-14

Browse, J., Somerville, C.R. (1991) Ann. Rev. Plant Physiol. Mol. Biol. 42:467-506 .

Chang, C., Bowman, J.L., DeJohn, A.W., Lander, E.S., Meyerowitz, E.M. (1988) Proc Natl Acad Sci USA 85:6856-6860

Christou, P., McCabe, D. E., Martinell, B. J., and Swain, W. F. (1990) Trends Biotechnol 8:145-151.

Church, G.M., and Gilbert, W. (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991-1995

Doyle, J. J., Schuler, M. A., Godette, W. D., Zenger, V., and Beachy, R. N. (1986) J Biol Chem 261:9228-9238.

Farmer, E. E., and Ryan, C. A. (1992) Plant Cell 4:129-134.

Fromm, M., (1990) UCLA Symposium on Molecular Strategies for Crop Improvement, April 16-22, 1990. Keystone, CO.

Grill, E., and C.R. Somerville (1991) Molec. Gen. Genet., 226:484-490.

Harwood, J. L. (1988) Annu Rev Plant Physiol. 39:101-38

Herrera-Estrella, L., et al. (1983) Nature 303:209

Hinchee, M.A.W., Connor-Ward, D.V., Newell, C.A., McDonnell, R.E., Sato, S. J., Gasser, C.S., Fischoff, D.A., Re, D.B., Fraley, R.T., Horsch, R.B. (1988) Bio/Technology 6:915-922.

Horsch, R.B. and H. Klee. (1986) Proc. Natl. Acad. Sci. U.S.A. 83:4428-32.

Hugly, S., Kunst, L., and Somerville, C. R. (1991) J. Hered. 82:484-488.

Hugly, S., and Somerville, C. R. (1992) Plant Physiol. 99:197-202.

Katagiri, F., E. Lam and N. Chua. (1989) Nature 340:727-730.

Kearns, E.V., S. Hugly, C.R. Somerville (1991) Arch. Biochem. Biophys., 284:431-436.

Klee, H.J., et al. (1985) Bio/Technology 3:637-42.

Knutson, D. S., Thompson, G. A., Radke, S. E., Johnson, W. B., Knauf, V. C., and Kridl, J. C. (1992) Proc. Natl. Acad. Sci. USA 89:2624-2628.

Lemieux, B. M. Miquel, C.R. Somerville, J. (1990) Theor. Appl. Genet., 80:234-240.

Lipman, D. J., and Pearson, W. R. (1985) Science 227:1435-1441.

McKeon T. A. , Stumpf, P. K. (1982) J. Biol. Chem. 257:12141-

47

McCourt, P., Kunst, L., Browse, J., Somerville, C. R. (1987) Plant Physiol. 84:353-361.

McSheffrey, S.A., McHughen A. and Devine, M.D., (1992) Theor. Appl. Genet., 84:480-486.

Meinke, D. W., Chen, J., and Beachy, R. N. (1981) Planta 153:130-139.

Mersereau, M., Pazour, J., and Das, A. (1990) Gene 90:149-151.

Rogers, S., et al. (1987) In 153 Methods in Enzymology. Edited by H. Weissbach and A. Weissbach. 253: Academic Press.

Samac, D.A., C.M. Hironaka, P.E. Yallaly and D.M. Shah (1990) Plant Physiol. 93:907-914.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Shanklin, J., C.R. Somerville. (1991) Proc. Natl. Acad. Sci. USA 88:2510-2514.

Schilperoot et al, EPO publication 120,516

Schmidhauser, T.J. and D.R. Helinski. (1985) J. Bacteriol. 164: 155.

Schmidt, H., Heinz, E. (1990a) Plant Physiol. 94:214-20

Schmidt, H., Heinz, E. (1990b) Proc. Natl.Acad. Sci. USA. 87:9477-9480.

Sherman, F., Fink, J., and Hicks, J. B. (1986) Laboratory Course Manual for Methods in Yeast Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shimamoto, K. et al. (1989) Nature 338:274-276.

Steponkus, P. L., Lynch, D. V., and Vemura, M. (1990) Phil. Trans. Roy. Soc. Lond. B. 326:571-583.

Sukumaran, N.P. and Weiser, C.J. (1972) HortScience 7:467-468.

Vasil, V., F. Redway and I. Vasil. (1990) Bio/Technology 8:429-434.

Wada, H., Gombos, Z., Murata, N. (1990) Nature 347:200-203.

Vollrath, D., Davis, R.W. (1987) Nucleic Acids Res 15:7865-7876.

Ward, E.R., Jen, G.C. (1990) Plant Mol Biol 14:561-568.

SEQUENCE LISTIN

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Monsanto Company
(B) STREET: 800 North Lindbergh Boulevard
(C) CITY: St. Louis
(D) STATE: Missouri
(E) COUNTRY: United States of America
(F) POSTAL CODE (ZIP): 63167
(G) TELEPHONE: (314)694-3131
(H) TELEFAX: (314)694-5435

(ii) TITLE OF INVENTION: Altered Linolenic and Linoleic Acid Content in Plants

(iii) NUMBER OF SEQUENCES: 72

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/156551
(B) FILING DATE: 22-NOV-1993

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/014431
(B) FILING DATE: 05-FEB-1993

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1353 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 87..1238

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AATCCATCAA ACCTTTATTC ACCACATTTC ACTGAAAGGC CACACATCTA GAGAGAGAAA 60

CTTCGTCCAA ATCTCTCTCT CCAGCG ATG GTT GTT GCT ATG GAC CAG CGC AGC 113

Met Val Val Ala Met Asp Gln Arg Ser
1 5

AAT GTT AAC GGA GAT TCC GGT GCC CGG AAG GAA GAA GGG TTT GAT CCA 161
Asn Val Asn Gly Asp Ser Gly Ala Arg Lys Glu Glu Gly Phe Asp Pro
10 15 20 25

AGC GCA CAA CCA CCG TTT AAG ATC GGA GAT ATA AGG GCG GCG ATT CCT 209
Ser Ala Gln Pro Pro Phe Lys Ile Gly Asp Ile Arg Ala Ala Ile Pro
30 35 40

AAG CAT TGC TGG GTG AAG AGT CCT TTG AGA TCT ATG AGC TAC GTC ACC 257
Lys His Cys Trp Val Lys Ser Pro Leu Arg Ser Met Ser Tyr Val Thr
45 50 55

AGA GAC ATT TTC GCC GTC GCG GCT CTG GCC ATG GCC GCC GTG TAT TTT 305
Arg Asp Ile Phe Ala Val Ala Ala Leu Ala Met Ala Ala Val Tyr Phe
60 65 70

GAT AGC TGG TTC CTC TGG CCA CTC TAC TGG GTT GCC CAA GGA ACC CTT 353
Asp Ser Trp Phe Leu Trp Pro Leu Tyr Trp Val Ala Gln Gly Thr Leu
75 80 85

TTC TGG GCC ATC TTC GTT CTT GGC CAC GAC TGT GGA CAT GGG AGT TTC 401
Phe Trp Ala Ile Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe
90 95 100 105

TCA GAC ATT CCT CTG CTG AAC AGT GTG GTT GGT CAC ATT CTT CAT TCA 449
Ser Asp Ile Pro Leu Leu Asn Ser Val Val Gly His Ile Leu His Ser
110 115 120

TTC ATC CTC GTT CCT TAC CAT GGT TGG AGA ATA AGC CAT CGG ACA CAC 497
Phe Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His Arg Thr His
125 130 135

CAC CAG AAC CAT GGC CAT GTT GAA AAC GAC GAG TCT TGG GTT CCG TTG 545
His Gln Asn His Gly His Val Glu Asn Asp Glu Ser Trp Val Pro Leu
140 145 150

CCA GAA AAG TTG TAC AAG AAC TTG CCC CAT AGT ACT CGG ATG CTC AGA 593
Pro Glu Lys Leu Tyr Lys Asn Leu Pro His Ser Thr Arg Met Leu Arg
155 160 165

TAC ACT GTC CCT CTG CCC ATG CTC GCT TAC CCG ATC TAT CTG TGG TAC 641
Tyr Thr Val Pro Leu Pro Met Leu Ala Tyr Pro Ile Tyr Leu Trp Tyr
170 175 180 185

AGA AGT CCT GGA AAA GAA GGG TCA CAT TTT AAC CCA TAC AGT AGT TTA 689
Arg Ser Pro Gly Lys Glu Gly Ser His Phe Asn Pro Tyr Ser Ser Leu
190 195 200

TTT GCT CCA AGC GAG AGG AAG CTT ATT GCA ACT TCA ACT ACT TGC TGG 737
Phe Ala Pro Ser Glu Arg Lys Leu Ile Ala Thr Ser Thr Thr Cys Trp
205 210 215

TCC ATA ATG TTG GCC ACT CTT GTT TAT CTA TCG TTC CTC GTT GAT CCA 785

```
Ser Ile Met Leu Ala Thr Leu Val Tyr Leu Ser Phe Leu Val Asp Pro
        220                 225                 230

GTC ACA GTT CTC AAA GTC TAT GGC GTT CCT TAC ATT ATC TTT GTG ATG      833
Val Thr Val Leu Lys Val Tyr Gly Val Pro Tyr Ile Ile Phe Val Met
    235                 240                 245

TGG TTG GAC GCT GTC ACG TAC TTG CAT CAT CAT GGT CAC GAT GAG AAG      881
Trp Leu Asp Ala Val Thr Tyr Leu His His His Gly His Asp Glu Lys
250                 255                 260                 265

TTG CCT TGG TAC AGA GGC AAG GAA TGG AGT TAT TTA CGT GGA GGA TTA      929
Leu Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu
                270                 275                 280

ACA ACT ATT GAT AGA GAT TAC GGA ATC TTC AAC AAC ATC CAT CAC GAC      977
Thr Thr Ile Asp Arg Asp Tyr Gly Ile Phe Asn Asn Ile His His Asp
            285                 290                 295

ATT GGA ACT CAC GTG ATC CAT CAT CTT TTC CCA CAA ATC CCT CAC TAT     1025
Ile Gly Thr His Val Ile His His Leu Phe Pro Gln Ile Pro His Tyr
        300                 305                 310

CAC TTG GTC GAT GCC ACG AGA GCA GCT AAA CAT GTG TTA GGA AGA TAC     1073
His Leu Val Asp Ala Thr Arg Ala Ala Lys His Val Leu Gly Arg Tyr
    315                 320                 325

TAC AGA GAG CCG AAG ACG TCA GGA GCA ATA CCG ATT CAC TTG GTG GAG     1121
Tyr Arg Glu Pro Lys Thr Ser Gly Ala Ile Pro Ile His Leu Val Glu
330                 335                 340                 345

AGT TTG GTC GCA AGT ATT AAA AAA GAT CAT TAC GTC AGT GAC ACT GGT     1169
Ser Leu Val Ala Ser Ile Lys Lys Asp His Tyr Val Ser Asp Thr Gly
                350                 355                 360

GAT ATT GTC TTC TAC GAG ACA GAT CCA GAT CTC TAC GTT TAT GCT TCT     1217
Asp Ile Val Phe Tyr Glu Thr Asp Pro Asp Leu Tyr Val Tyr Ala Ser
            365                 370                 375

GAC AAA TCT AAA ATC AAT TAACTTTTCT TCCTAGCTCT ATTAGGAATA            1265
Asp Lys Ser Lys Ile Asn
        380

AACACTCCTT CTCTTTTACT TATTTGTTTC TGCTTTAAGT TTAAAATGTA CTCGTGAAAC   1325

CTTTTTTTTA TTAATGTATT TACGTTAC                                      1353
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 383 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTI N: SEQ ID NO:2:

Met Val Val Ala Met Asp Gln Arg Ser Asn Val Asn Gly Asp Ser Gly
1 5 10 15

Ala Arg Lys Glu Glu Gly Phe Asp Pro Ser Ala Gln Pro Pro Phe Lys
20 25 30

Ile Gly Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp Val Lys Ser
35 40 45

Pro Leu Arg Ser Met Ser Tyr Val Thr Arg Asp Ile Phe Ala Val Ala
50 55 60

Ala Leu Ala Met Ala Ala Val Tyr Phe Asp Ser Trp Phe Leu Trp Pro
65 70 75 80

Leu Tyr Trp Val Ala Gln Gly Thr Leu Phe Trp Ala Ile Phe Val Leu
85 90 95

Gly His Asp Cys Gly His Gly Ser Phe Ser Asp Ile Pro Leu Leu Asn
100 105 110

Ser Val Val Gly His Ile Leu His Ser Phe Ile Leu Val Pro Tyr His
115 120 125

Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His Gly His Val
130 135 140

Glu Asn Asp Glu Ser Trp Val Pro Leu Pro Glu Lys Leu Tyr Lys Asn
145 150 155 160

Leu Pro His Ser Thr Arg Met Leu Arg Tyr Thr Val Pro Leu Pro Met
165 170 175

Leu Ala Tyr Pro Ile Tyr Leu Trp Tyr Arg Ser Pro Gly Lys Glu Gly
180 185 190

Ser His Phe Asn Pro Tyr Ser Ser Leu Phe Ala Pro Ser Glu Arg Lys
195 200 205

Leu Ile Ala Thr Ser Thr Thr Cys Trp Ser Ile Met Leu Ala Thr Leu
210 215 220

Val Tyr Leu Ser Phe Leu Val Asp Pro Val Thr Val Leu Lys Val Tyr
225 230 235 240

Gly Val Pro Tyr Ile Ile Phe Val Met Trp Leu Asp Ala Val Thr Tyr
245 250 255

Leu His His His Gly His Asp Glu Lys Leu Pro Trp Tyr Arg Gly Lys
260 265 270

Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Ile Asp Arg Asp Tyr
275 280 285

```
Gly Ile Phe Asn Asn Ile His His Asp Ile Gly Thr His Val Ile His
    290                 295                 300

His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Asp Ala Thr Arg
305                 310                 315                 320

Ala Ala Lys His Val Leu Gly Arg Tyr Tyr Arg Glu Pro Lys Thr Ser
                325                 330                 335

Gly Ala Ile Pro Ile His Leu Val Glu Ser Leu Val Ala Ser Ile Lys
            340                 345                 350

Lys Asp His Tyr Val Ser Asp Thr Gly Asp Ile Val Phe Tyr Glu Thr
        355                 360                 365

Asp Pro Asp Leu Tyr Val Tyr Ala Ser Asp Lys Ser Lys Ile Asn
    370                 375                 380
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGCGATGCTG TCGGAATGGA CGATA 25

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CTTGGAGCCA CTATCGACTA CGCGATC 27

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CCGATCTCAA GATTACGGAA T 21

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TTCCTAATGC AGGAGTCGCA TAAG 24

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AGGAGTCGCA TAAGGGAG 18

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGGAAGTGAA TGGAGAC 17

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1645 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 125..1465

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
GGAAAACACA AGTTTCTCTC ACACACATTA TCTCTTTCTC TATTACCACC ACTCATTCAT      60

AACAGAAACC CACCAAAAAA TAAAAAGAGA GACTTTTCAC TCTGGGGAGA GAGCTCAAGT     120

TCTA ATG GCG AAC TTG GTC TTA TCA GAA TGT GGT ATA CGA CCT CTC CCC      169
     Met Ala Asn Leu Val Leu Ser Glu Cys Gly Ile Arg Pro Leu Pro
      1               5                  10                  15

AGA ATC TAC ACA ACA CCC AGA TCC AAT TTC CTC TCC AAC AAC AAC AAA       217
Arg Ile Tyr Thr Thr Pro Arg Ser Asn Phe Leu Ser Asn Asn Asn Lys
                20                  25                  30

TTC AGA CCA TCA CTT TCT TCT TCT TCT TAC AAA ACA TCA TCA TCT CCT       265
Phe Arg Pro Ser Leu Ser Ser Ser Ser Tyr Lys Thr Ser Ser Ser Pro
            35                  40                  45

CTG TCT TTT GGT CTG AAT TCA CGA GAT GGG TTC ACG AGG AAT TGG GCG       313
Leu Ser Phe Gly Leu Asn Ser Arg Asp Gly Phe Thr Arg Asn Trp Ala
        50                  55                  60

TTG AAT GTG AGC ACA CCA TTA ACG ACA CCA ATA TTT GAG GAG TCT CCA       361
Leu Asn Val Ser Thr Pro Leu Thr Thr Pro Ile Phe Glu Glu Ser Pro
    65                  70                  75

TTG GAG GAA GAT AAT AAA CAG AGA TTC GAT CCA GGT GCG CCT CCT CCG       409
Leu Glu Glu Asp Asn Lys Gln Arg Phe Asp Pro Gly Ala Pro Pro Pro
80                  85                  90                  95

TTC AAT TTA GCT GAT ATT AGA GCA GCT ATA CCT AAG CAT TGT TGG GTT       457
Phe Asn Leu Ala Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp Val
                100                 105                 110

AAG AAT CCA TGG AAG TCT TTG AGT TAT GTC GTC AGA GAC GTC GCT ATC       505
Lys Asn Pro Trp Lys Ser Leu Ser Tyr Val Val Arg Asp Val Ala Ile
            115                 120                 125
```

```
GTC TTT GCA TTG GCT GCT GGA GCT GCT TAC CTC AAC AAT TGG ATT GTT    553
Val Ph  Ala Leu Ala Ala Gly Ala Ala Tyr Leu Asn Asn Trp Ile Val
        130                 135                 140

TGG CCT CTC TAT TGG CTC GCT CAA GGA ACC ATG TTT TGG GCT CTC TTT    601
Trp Pro Leu Tyr Trp Leu Ala Gln Gly Thr Met Phe Trp Ala Leu Phe
    145                 150                 155

GTT CTT GGT CAT GAC TGT GGA CAT GGT AGT TTC TCA AAT GAT CCG AAG    649
Val Leu Gly His Asp Cys Gly His Gly Ser Phe Ser Asn Asp Pro Lys
160                 165                 170                 175

TTG AAC AGT GTG GTC GGT CAT CTT CTT CAT TCC TCA ATT CTG GTC CCA    697
Leu Asn Ser Val Val Gly His Leu Leu His Ser Ser Ile Leu Val Pro
                180                 185                 190

TAC CAT GGC TGG AGA ATT AGT CAC AGA ACT CAC CAC CAG AAC CAT GGA    745
Tyr His Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His Gly
            195                 200                 205

CAT GTT GAG AAT GAC GAA TCT TGG CAT CCT ATG TCT GAG AAA ATC TAC    793
His Val Glu Asn Asp Glu Ser Trp His Pro Met Ser Glu Lys Ile Tyr
        210                 215                 220

AAT ACT TTG GAC AAG CCG ACT AGA TTC TTT AGA TTT ACA CTG CCT CTC    841
Asn Thr Leu Asp Lys Pro Thr Arg Phe Phe Arg Phe Thr Leu Pro Leu
    225                 230                 235

GTG ATG CTT GCA TAC CCT TTC TAC TTG TGG GCT CGA AGT CCG GGG AAA    889
Val Met Leu Ala Tyr Pro Phe Tyr Leu Trp Ala Arg Ser Pro Gly Lys
240                 245                 250                 255

AAG GGT TCT CAT TAC CAT CCA GAC AGT GAC TTG TTC CTC CCT AAA GAG    937
Lys Gly Ser His Tyr His Pro Asp Ser Asp Leu Phe Leu Pro Lys Glu
                260                 265                 270

AGA AAG GAT GTC CTC ACT TCT ACT GCT TGT TGG ACT GCA ATG GCT GCT    985
Arg Lys Asp Val Leu Thr Ser Thr Ala Cys Trp Thr Ala Met Ala Ala
            275                 280                 285

CTG CTT GTT TGT CTC AAC TTC ACA ATC GGT CCA ATT CAA ATG CTC AAA   1033
Leu Leu Val Cys Leu Asn Phe Thr Ile Gly Pro Ile Gln Met Leu Lys
        290                 295                 300

CTT TAT GGA ATT CCT TAC TGG ATA AAT GTA ATG TGG TTG GAC TTT GTG   1081
Leu Tyr Gly Ile Pro Tyr Trp Ile Asn Val Met Trp Leu Asp Phe Val
    305                 310                 315

ACT TAC CTG CAT CAC CAT GGT CAT GAA GAT AAG CTT CCT TGG TAC CGT   1129
Thr Tyr Leu His His His Gly His Glu Asp Lys Leu Pro Trp Tyr Arg
320                 325                 330                 335

GGC AAG GAG TGG AGT TAC CTG AGA GGA GGA CTT ACA ACA TTG GAT CGT   1177
Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Leu Asp Arg
                340                 345                 350
```

GAC TAC GGA TTG ATC AAT AAC ATC CAT CAT GAT ATT GGA ACT CAT GTG 1225
Asp Tyr Gly Leu Ile Asn Asn Ile His His Asp Ile Gly Thr His Val
355 360 365

ATA CAT CAT CTT TTC CCG CAG ATC CCA CAT TAT CAT CTA GTA GAA GCA 1273
Ile His His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Glu Ala
370 375 380

ACA GAA GCA GCT AAA CCA GTA TTA GGG AAG TAT TAC AGG GAG CCT GAT 1321
Thr Glu Ala Ala Lys Pro Val Leu Gly Lys Tyr Tyr Arg Glu Pro Asp
385 390 395

AAG TCT GGA CCG TTG CCA TTA CAT TTA CTG GAA ATT CTA GCG AAA AGT 1369
Lys Ser Gly Pro Leu Pro Leu His Leu Leu Glu Ile Leu Ala Lys Ser
400 405 410 415

ATA AAA GAA GAT CAT TAC GTG AGC GAC GAA GGA GAA GTT GTA TAC TAT 1417
Ile Lys Glu Asp His Tyr Val Ser Asp Glu Gly Glu Val Val Tyr Tyr
420 425 430

AAA GCA GAT CCA AAT CTC TAT GGA GAG GTC AAA GTA AGA GCA GAT TGAAATGAAG
1472
Lys Ala Asp Pro Asn Leu Tyr Gly Glu Val Lys Val Arg Ala Asp
435 440 445

CAGGCTTGAG ATTGAAGTTT TTTCTATTTC AGACCAGCTG ATTTTTTGCT TACTGTATCA 1532

ATTTATTGTG TCACCCACCA GAGAGTTAGT ATCTCTGAAT ACGATCGATC AGATGGAAAC 1592

AACAAATTTG TTTGCGATAC TGAAGCTATA TATACCATAA AAAAAAAAAA AAA 1645

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 446 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Ala Asn Leu Val Leu Ser Glu Cys Gly Ile Arg Pro Leu Pro Arg
1 5 10 15

Ile Tyr Thr Thr Pro Arg Ser Asn Phe Leu Ser Asn Asn Asn Lys Phe
20 25 30

Arg Pro Ser Leu Ser Ser Ser Ser Tyr Lys Thr Ser Ser Ser Pro Leu
35 40 45

Ser Phe Gly Leu Asn Ser Arg Asp Gly Phe Thr Arg Asn Trp Ala Leu
50 55 60

Asn Val Ser Thr Pro Leu Thr Thr Pro Ile Phe Glu Glu Ser Pro Leu

65 70 75 80

Glu Glu Asp Asn Lys Gln Arg Phe Asp Pro Gly Ala Pr Pro Pro Phe
85 90 95

Asn Leu Ala Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp Val Lys
100 105 110

Asn Pro Trp Lys Ser Leu Ser Tyr Val Val Arg Asp Val Ala Ile Val
115 120 125

Phe Ala Leu Ala Ala Gly Ala Ala Tyr Leu Asn Asn Trp Ile Val Trp
130 135 140

Pro Leu Tyr Trp Leu Ala Gln Gly Thr Met Phe Trp Ala Leu Phe Val
145 150 155 160

Leu Gly His Asp Cys Gly His Gly Ser Phe Ser Asn Asp Pro Lys Leu
165 170 175

Asn Ser Val Val Gly His Leu Leu His Ser Ser Ile Leu Val Pro Tyr
180 185 190

His Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His Gly His
195 200 205

Val Glu Asn Asp Glu Ser Trp His Pro Met Ser Glu Lys Ile Tyr Asn
210 215 220

Thr Leu Asp Lys Pro Thr Arg Phe Phe Arg Phe Thr Leu Pro Leu Val
225 230 235 240

Met Leu Ala Tyr Pro Phe Tyr Leu Trp Ala Arg Ser Pro Gly Lys Lys
245 250 255

Gly Ser His Tyr His Pro Asp Ser Asp Leu Phe Leu Pro Lys Glu Arg
260 265 270

Lys Asp Val Leu Thr Ser Thr Ala Cys Trp Thr Ala Met Ala Ala Leu
275 280 285

Leu Val Cys Leu Asn Phe Thr Ile Gly Pro Ile Gln Met Leu Lys Leu
290 295 300

Tyr Gly Ile Pro Tyr Trp Ile Asn Val Met Trp Leu Asp Phe Val Thr
305 310 315 320

Tyr Leu His His His Gly His Glu Asp Lys Leu Pro Trp Tyr Arg Gly
325 330 335

Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Leu Asp Arg Asp
340 345 350

Tyr Gly Leu Ile Asn Asn Ile His His Asp Ile Gly Thr His Val Ile
355 360 365

His His Leu Ph Pro Gln Il Pro His Tyr His Leu Val Glu Ala Thr
370 375 380

Glu Ala Ala Lys Pro Val Leu Gly Lys Tyr Tyr Arg Glu Pro Asp Lys
385 390 395 400

Ser Gly Pro Leu Pro Leu His Leu Leu Glu Ile Leu Ala Lys Ser Ile
405 410 415

Lys Glu Asp His Tyr Val Ser Asp Glu Gly Glu Val Val Tyr Tyr Lys
420 425 430

Ala Asp Pro Asn Leu Tyr Gly Glu Val Lys Val Arg Ala Asp
435 440 445

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1525 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 61..1368

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AGAGAGTGCA AATAGAACGA CAGAGACTTT TTCCTCTTTT CTTCTTGGGA AGAGGCTCCA 60

ATG GCG AGC TCG GTT TTA TCA GAA TGT GGT TTT AGA CCT CTC CCC AGA 108
Met Ala Ser Ser Val Leu Ser Glu Cys Gly Phe Arg Pro Leu Pro Arg
1 5 10 15

TTC TAC CCT AAA CAC ACA ACC TCT TTT GCC TCT AAC CCT AAA CCC ACT 156
Phe Tyr Pro Lys His Thr Thr Ser Phe Ala Ser Asn Pro Lys Pro Thr
20 25 30

TTC AAA TTC AAT CCA CCA CTT AAA CCT CCT TCT TCT CTT CTC AAT TCC 204
Phe Lys Phe Asn Pro Pro Leu Lys Pro Pro Ser Ser Leu Leu Asn Ser
35 40 45

CGA TAT GGA TTC TAC TCT AAA ACC AGG AAC TGG GCA TTG AAT GTG GCA 252
Arg Tyr Gly Phe Tyr Ser Lys Thr Arg Asn Trp Ala Leu Asn Val Ala
50 55 60

ACA CCT TTA ACA ACT CTT CAG TCT CCA TCC GAG GAA GAC ACG GAG AGA 300
Thr Pro Leu Thr Thr Leu Gln Ser Pro Ser Glu Glu Asp Thr Glu Arg
65 70 75 80

```
TTC GAC CCA GGT GCG CCT CCT CCC TTC AAT TTG GCG GAT ATA AGA GCA    348
Phe Asp Pr  Gly Ala Pro Pro Pro Phe Asn Leu Ala Asp Ile Arg Ala
                85                  90                  95

GCC ATA CCT AAG CAT TGT TGG GTT AAG AAT CCA TGG ATG TCT ATG AGT    396
Ala Ile Pro Lys His Cys Trp Val Lys Asn Pro Trp Met Ser Met Ser
            100                 105                 110

TAT GTT GTC AGA GAT GTT GCT ATC GTC TTT GGA TTG GCT GCT GTT GCT    444
Tyr Val Val Arg Asp Val Ala Ile Val Phe Gly Leu Ala Ala Val Ala
        115                 120                 125

GCT TAC TTC AAC AAT TGG CTT CTC TGG CCT CTC TAC TGG TTC GCT CAA    492
Ala Tyr Phe Asn Asn Trp Leu Leu Trp Pro Leu Tyr Trp Phe Ala Gln
    130                 135                 140

GGA ACC ATG TTC TGG GCT CTC TTT GTC CTT GGC CAT GAC TGC GGA CAT    540
Gly Thr Met Phe Trp Ala Leu Phe Val Leu Gly His Asp Cys Gly His
145                 150                 155                 160

GGT AGC TTC TCG AAT GAT CCG AGG CTG AAC AGT GTG GCT GGT CAT CTT    588
Gly Ser Phe Ser Asn Asp Pro Arg Leu Asn Ser Val Ala Gly His Leu
                165                 170                 175

CTT CAT TCC TCA ATT CTG GTC CCT TAC CAT GGC TGG AGG ATT AGC CAC    636
Leu His Ser Ser Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His
            180                 185                 190

AGA ACT CAC CAC CAG AAC CAT GGT CAT GTC GAG AAT GAC GAA TCA TGG    684
Arg Thr His His Gln Asn His Gly His Val Glu Asn Asp Glu Ser Trp
        195                 200                 205

CAT CCT TTG CCT GAA AGC ATC TAC AAG AAT TTG GAA AAG ACG ACT CAA    732
His Pro Leu Pro Glu Ser Ile Tyr Lys Asn Leu Glu Lys Thr Thr Gln
    210                 215                 220

ATG TTT AGG TTT ACA CTG CCT TTT CCA ATG CTC GCA TAC CCT TTC TAC    780
Met Phe Arg Phe Thr Leu Pro Phe Pro Met Leu Ala Tyr Pro Phe Tyr
225                 230                 235                 240

TTG TGG AAC AGA AGT CCA GGG AAA CAA GGT TCT CAT TAT CAT CCG GAC    828
Leu Trp Asn Arg Ser Pro Gly Lys Gln Gly Ser His Tyr His Pro Asp
                245                 250                 255

AGT GAC TTG TTT CTT CCA AAA GAG AAG AAA GAT GTT CTG ACA TCA ACT    876
Ser Asp Leu Phe Leu Pro Lys Glu Lys Lys Asp Val Leu Thr Ser Thr
            260                 265                 270

GCC TGT TGG ACT GCA ATG GCT GCT TTG CTT GTT TGT CTC AAC TTT GTC    924
Ala Cys Trp Thr Ala Met Ala Ala Leu Leu Val Cys Leu Asn Phe Val
        275                 280                 285

ATG GGT CCA ATC CAG ATG CTC AAA CTA TAT GGC ATC CCT TAT TGG ATA    972
Met Gly Pro Ile Gln Met Leu Lys Leu Tyr Gly Ile Pro Tyr Trp Ile
    290                 295                 300
```

TTT GTA ATG TGG TTG GAC TTC GTC ACT TAC TTG CAC CAC CAT GGA CAT 1020
Phe Val Met Trp Leu Asp Phe Val Thr Tyr Leu His His His Gly His
305 310 315 320

GAA GAC AAG CTC CCT TGG TAT CGT GGA AAG GAA TGG AGT TAC CTG AGA 1068
Glu Asp Lys Leu Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg
325 330 335

GGA GGG CTC ACA ACA TTA GAT CGT GAC TAC GGA TGG ATC AAT AAC ATC 1116
Gly Gly Leu Thr Thr Leu Asp Arg Asp Tyr Gly Trp Ile Asn Asn Ile
340 345 350

CAC CAC GAT ATT GGA ACT CAT GTG ATA CAT CAT CTT TTC CCG CAG ATC 1164
His His Asp Ile Gly Thr His Val Ile His His Leu Phe Pro Gln Ile
355 360 365

CCA CAT TAT CAT CTA GTA GAA GCA ACA GAA GCA GCT AAA CCA GTA CTA 1212
Pro His Tyr His Leu Val Glu Ala Thr Glu Ala Ala Lys Pro Val Leu
370 375 380

GGA AAG TAC TAC AGA GAA CCG AAA AAC TCT GGA CCT CTG CCA CTT CAC 1260
Gly Lys Tyr Tyr Arg Glu Pro Lys Asn Ser Gly Pro Leu Pro Leu His
385 390 395 400

TTA CTG GGA AGC CTC ATA AAG AGT ATG AAA CAA GAC CAT TTC GTA AGC 1308
Leu Leu Gly Ser Leu Ile Lys Ser Met Lys Gln Asp His Phe Val Ser
405 410 415

GAT ACA GGA GAT GTC GTG TAC TAT GAG GCA GAT CCA AAA CTC AAT GGA 1356
Asp Thr Gly Asp Val Val Tyr Tyr Glu Ala Asp Pro Lys Leu Asn Gly
420 425 430

CAA AGA ACA TGAGGACATA CTGCAGTGAA CCAGGCAGAC AAGTTACATA 1405
Gln Arg Thr
435

AATTCATCTT GGCCCATTCA TTATGTTCTT TTTGTTTTGG TGTAAAGCCT TTTCGAGATT 1465

AAAAAAGCAT TAATTTGTAG AAACCTGTGG TAAAACTCTC GATCAAATGA AATAAGATAT 1525

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 435 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Ala Ser Ser Val Leu Ser Glu Cys Gly Phe Arg Pro Leu Pro Arg
1 5 10 15

Phe Tyr Pro Lys His Thr Thr Ser Phe Ala Ser Asn Pr Lys Pr Thr
            20   25   30

Phe Lys Phe Asn Pro Pro Leu Lys Pro Pro Ser Ser Leu Leu Asn Ser
        35   40   45

Arg Tyr Gly Phe Tyr Ser Lys Thr Arg Asn Trp Ala Leu Asn Val Ala
    50   55   60

Thr Pro Leu Thr Thr Leu Gln Ser Pro Ser Glu Glu Asp Thr Glu Arg
65   70   75   80

Phe Asp Pro Gly Ala Pro Pro Pro Phe Asn Leu Ala Asp Ile Arg Ala
                85   90   95

Ala Ile Pro Lys His Cys Trp Val Lys Asn Pro Trp Met Ser Met Ser
            100   105   110

Tyr Val Val Arg Asp Val Ala Ile Val Phe Gly Leu Ala Ala Val Ala
        115   120   125

Ala Tyr Phe Asn Asn Trp Leu Leu Trp Pro Leu Tyr Trp Phe Ala Gln
    130   135   140

Gly Thr Met Phe Trp Ala Leu Phe Val Leu Gly His Asp Cys Gly His
145   150   155   160

Gly Ser Phe Ser Asn Asp Pro Arg Leu Asn Ser Val Ala Gly His Leu
                165   170   175

Leu His Ser Ser Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His
            180   185   190

Arg Thr His His Gln Asn His Gly His Val Glu Asn Asp Glu Ser Trp
        195   200   205

His Pro Leu Pro Glu Ser Ile Tyr Lys Asn Leu Glu Lys Thr Thr Gln
    210   215   220

Met Phe Arg Phe Thr Leu Pro Phe Pro Met Leu Ala Tyr Pro Phe Tyr
225   230   235   240

Leu Trp Asn Arg Ser Pro Gly Lys Gln Gly Ser His Tyr His Pro Asp
                245   250   255

Ser Asp Leu Phe Leu Pro Lys Glu Lys Lys Asp Val Leu Thr Ser Thr
            260   265   270

Ala Cys Trp Thr Ala Met Ala Ala Leu Leu Val Cys Leu Asn Phe Val
        275   280   285

Met Gly Pro Ile Gln Met Leu Lys Leu Tyr Gly Ile Pro Tyr Trp Ile
    290   295   300

Phe Val Met Trp Leu Asp Phe Val Thr Tyr Leu His His His Gly His
305   310   315   320

```
Glu Asp Lys Leu Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg
                325                 330                 335

Gly Gly Leu Thr Thr Leu Asp Arg Asp Tyr Gly Trp Ile Asn Asn Ile
            340                 345                 350

His His Asp Ile Gly Thr His Val Ile His His Leu Phe Pro Gln Ile
        355                 360                 365

Pro His Tyr His Leu Val Glu Ala Thr Glu Ala Ala Lys Pro Val Leu
    370                 375                 380

Gly Lys Tyr Tyr Arg Glu Pro Lys Asn Ser Gly Pro Leu Pro Leu His
385                 390                 395                 400

Leu Leu Gly Ser Leu Ile Lys Ser Met Lys Gln Asp His Phe Val Ser
                405                 410                 415

Asp Thr Gly Asp Val Val Tyr Tyr Glu Ala Asp Pro Lys Leu Asn Gly
            420                 425                 430

Gln Arg Thr
        435
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAYATHMGNG CNGCNATHCC 20

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCNATHCCNA ARCAYTG 17

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AARCAYTGYT GGGTNAA 17

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TGGTTYYTNT GGCCNYTNTA YTGG 24

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TGGTTYYTNT GGCCN 15

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synth tic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TGGCCNYTNT AYTGG 15

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TGGGTNGCNC ARGGNAC 17

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TTYGTNYTNG GNCAYGA 17

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GTNYTNGGNC AYGAYTG 17

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GGNCAYGAYT GYGGNCA 17

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TGYGGNCAYG GNWSNTT 17

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCNTAYCAYG GNTGG 15

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CAYGGNTGGM GNATHWSNCA 20

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TGGMGNATHT CNCAYMGNAC NCAYCA 26

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TGGMGNATHA GYCAYMGNAC NCAYCA 26

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

TGGMGNATHW SNCAY 15

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CAYMGNACNC AYCAY 15

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GARAAYGAYG ARWSNTGG 18

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GAYGARWSNT GGGTNCC 17

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

NGTNACNGCR TCNARCCA 18

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

RTGRTGNARR TANGT 15

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

ARNCCNCCNC KNARRTARCT CCA 23

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

ARNCCNCCNC KNARRTANGA CCA 23

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

RTCNCKRTCD ATNGTNGTNA 20

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

RTARTCNCKR TCDATNGT 18

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

RTGNGTNCCD ATRTCRTG 18

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

NARRTGRTGD ATNACRTG 18

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

DATYTGNGGR AANARRTGRT G 21

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GGDATYTGNG GRAANARRTG 20

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (synthetic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

RTARTGNGGD ATYTGNGGRA ANA 23

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Asp Ile Arg Ala Ala Ile Pro
1 5

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Ala Ile Pro Lys His Cys
1 5

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Lys His Cys Trp Val Lys
1 5

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Trp Phe Leu Trp Pro Leu Tyr Trp
1 5

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Trp Phe Leu Trp Pro
1 5

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Trp Pro Leu Tyr Trp
1 5

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Trp Val Ala Gln Gly Thr
1 5

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Trp Val Ala Gln Gly Thr
1 5

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Val Leu Gly His Asp Cys
1 5

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Gly His Asp Cys Gly His
1 5

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Cys Gly His Gly Ser Phe
1 5

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Pro Tyr His Gly Trp
1 5

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

His Gly Trp Arg Ile Ser His
1 5

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Trp Arg Ile Ser His Arg Thr His His
1 5

(2) INFORMATION F R SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Trp Arg Ile Ser His
1 5

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

His Arg Thr His His
1 5

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Glu Asn Asp Glu Ser Trp
1 5

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Asp Glu Ser Trp Val Pro
1 5

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Trp Leu Asp Ala Val Thr
1 5

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Thr Tyr Leu His His
1 5

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Trp Ser Tyr Leu Arg Gly Gly Leu
1 5

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Leu Thr Thr Ile Asp Arg Asp
1 5

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Thr Ile Asp Arg Asp Tyr
1 5

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

His Asp Ile Gly Thr His
1 5

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

His Val Ile His His Leu
1 5

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

His His Leu Phe Pro Gln Ile
1 5

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

His Leu Phe Pro Gln Ile Pro
1 5

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

Leu Phe Pro Gln Ile Pro His Tyr
1 5

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1670 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 46..1302

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

```
CAAACTCTCT CGGGGGGTCG CTTCTTCTGC ATTTTCTGCT TCCCA ATG GCT TCC      54
                                                  Met Ala Ser
                                                   1

AGA ATT GCT GAT TCT CTC TTC GCC TTC ACG GGC CCA CAG CAA TGT CTT   102
Arg Ile Ala Asp Ser Leu Phe Ala Phe Thr Gly Pro Gln Gln Cys Leu
      5                  10                  15

CCT AGG GTT CCT AAG CTT GCT GCT TCT TCT GCT CGT GTT TCT CCT GGT   150
Pro Arg Val Pro Lys Leu Ala Ala Ser Ser Ala Arg Val Ser Pro Gly
 20                  25                  30                  35

GTA TAT GCT GTG AAG CCG ATT GAT CTT CTG TTA AAA GGA CGA ACT CAT   198
Val Tyr Ala Val Lys Pro Ile Asp Leu Leu Leu Lys Gly Arg Thr His
                 40                  45                  50

CGA AGT AGA AGA TGT GTA GCT CCT GTG AAA AGG AGA ATT GGA TGT ATC   246
Arg Ser Arg Arg Cys Val Ala Pro Val Lys Arg Arg Ile Gly Cys Ile
             55                  60                  65

AAA GCG GTG GCT GCT CCA GTT GCA CCG CCT TCA GCT GAC AGT GCA GAA   294
Lys Ala Val Ala Ala Pro Val Ala Pro Pro Ser Ala Asp Ser Ala Glu
         70                  75                  80

GAC AGG GAA CAG TTA GCA GAA AGC TAT GGA TTC AGA CAA ATT GGA GAA   342
Asp Arg Glu Gln Leu Ala Glu Ser Tyr Gly Phe Arg Gln Ile Gly Glu
     85                  90                  95

GAT CTT CCT GAG AAT GTC ACC TTA AAA GAT ATC ATG GAT ACA CTT CCC   390
Asp Leu Pro Glu Asn Val Thr Leu Lys Asp Ile Met Asp Thr Leu Pro
100                 105                 110                 115

AAA GAG GTG TTT GAG ATT GAT GAT CTG AAA GCT TTG AAG TCT GTG TTG   438
Lys Glu Val Phe Glu Ile Asp Asp Leu Lys Ala Leu Lys Ser Val Leu
                120                 125                 130

ATA TCT GTG ACT TCA TAC ACT TTG GGG CTC TTC ATG ATT GCA AAA TCG   486
Ile Ser Val Thr Ser Tyr Thr Leu Gly Leu Phe Met Ile Ala Lys Ser
            135                 140                 145

CCG TGG TAT CTG CTA CCG TTG GCT TGG GCA TGG ACA GGA ACT GCA ATT   534
```

Pro Trp Tyr Leu Leu Pro Leu Ala Trp Ala Trp Thr Gly Thr Ala Ile
150 155 160

ACC GGG TTC TTT GTG ATA GGT CAT GAT TGT GCA CAT AAG TCA TTT TCA 582
Thr Gly Phe Phe Val Ile Gly His Asp Cys Ala His Lys Ser Phe Ser
165 170 175

AAG AAC AAA TTG GTG GAA GAC ATT GTG GGT ACT CTC GCC TTC CTA CCA 630
Lys Asn Lys Leu Val Glu Asp Ile Val Gly Thr Leu Ala Phe Leu Pro
180 185 190 195

CTT GTC TAC CCA TAT GAG CCA TGG CGG TTT AAG CAC GAC CGC CAT CAC 678
Leu Val Tyr Pro Tyr Glu Pro Trp Arg Phe Lys His Asp Arg His His
200 205 210

GCC AAA ACC AAC ATG TTA CTT CAT GAC ACA GCT TGG CAG CCA GTT CCG 726
Ala Lys Thr Asn Met Leu Leu His Asp Thr Ala Trp Gln Pro Val Pro
215 220 225

CCA GAG GAG TTT GAG TCA TCA CCC GTG ATG AGA AAG GCA ATC ATT TTT 774
Pro Glu Glu Phe Glu Ser Ser Pro Val Met Arg Lys Ala Ile Ile Phe
230 235 240

GGA TAT GGC CCA ATT AGA CCT TGG TTG TCC ATA GCT CAC TGG GTG AAC 822
Gly Tyr Gly Pro Ile Arg Pro Trp Leu Ser Ile Ala His Trp Val Asn
245 250 255

TGG CAC TTC AAT CTG AAA AAG TTC AGA GCG AGC GAG GTG AAT AGG GTG 870
Trp His Phe Asn Leu Lys Lys Phe Arg Ala Ser Glu Val Asn Arg Val
260 265 270 275

AAG ATA AGT TTG GCT TGT GTT TTC GCC TTC ATG GCC GTT GGG TGG CCA 918
Lys Ile Ser Leu Ala Cys Val Phe Ala Phe Met Ala Val Gly Trp Pro
280 285 290

CTG ATC GTA TAC AAA GTT GGT ATA TTG GGA TGG GTA AAA TTC TGG TTA 966
Leu Ile Val Tyr Lys Val Gly Ile Leu Gly Trp Val Lys Phe Trp Leu
295 300 305

ATG CCA TGG TTG GGC TAT CAC TTC TGG ATG AGC ACA TTC ACA ATG GTT 1014
Met Pro Trp Leu Gly Tyr His Phe Trp Met Ser Thr Phe Thr Met Val
310 315 320

CAT CAT ACG GCT CCG CAT ATA CCT TTC AAG CCT GCG GAT GAG TGG AAC 1062
His His Thr Ala Pro His Ile Pro Phe Lys Pro Ala Asp Glu Trp Asn
325 330 335

GCG GCT CAG GCC CAG CTG AAT GGA ACT GTT CAT TGT GAC TAC CCT AGT 1110
Ala Ala Gln Ala Gln Leu Asn Gly Thr Val His Cys Asp Tyr Pro Ser
340 345 350 355

TGG ATT GAA ATT CTC TGC CAT GAT ATC AAC GTT CAC ATC CCG CAT CAT 1158
Trp Ile Glu Ile Leu Cys His Asp Ile Asn Val His Ile Pro His His
360 365 370

ATT AGC CCA AGA ATA CCG AGC TAC AAT CTC CGT GCA GCT CAT GAG TCT 1206

```
Ile Ser Pro Arg Ile Pro Ser Tyr Asn Leu Arg Ala Ala His Glu Ser
            375                 380                 385

ATA CAA GAG AAC TGG GGA AAG TAT ACA AAC TTG GCT ACA TGG AAC TGG      1254
Ile Gln Glu Asn Trp Gly Lys Tyr Thr Asn Leu Ala Thr Trp Asn Trp
        390                 395                 400

CGA TTG ATG AAG ACG ATA ATG ACT GTG TGT CAT GTC TAT GAC AAA TAGGAGAACT
1309
Arg Leu Met Lys Thr Ile Met Thr Val Cys His Val Tyr Asp Lys
    405                 410                 415

ACATTCCTTT TGACCGGTTA GCCCCTGAAG AATCTCAGCC AATAACCTTC CTCAAGAAAT     1369

CAATGCCTAA CTACACAGCC TGATTCGCCA TGGTCTCAAA CTAGTCTTTT GAAATCTCAA     1429

TATCTTTTTG CAGTCGCCGA TGTTATATGT AAGCTTTCCA AGCGATGAGC TTCTCTAACA     1489

CTTCACCAAC GCTTTATACT GTTATCTTCT TTCCAATCTT ATCAGAAGAG AGAAACTGGT     1549

CAAATTATCT GAGCGATTGC AATTCTTTTA TCAGTTTCTT AGCTATAAGA AGATTGAACA     1609

GTCTATATAG TTTGCAATGT ACTGTAATGT GATGAAAATT TAGTTGATGA GAAAAAAAAA     1669

A                                                                    1670
```

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 418 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

```
Met Ala Ser Arg Ile Ala Asp Ser Leu Phe Ala Phe Thr Gly Pro Gln
  1               5                  10                  15

Gln Cys Leu Pro Arg Val Pro Lys Leu Ala Ala Ser Ser Ala Arg Val
             20                  25                  30

Ser Pro Gly Val Tyr Ala Val Lys Pro Ile Asp Leu Leu Leu Lys Gly
         35                  40                  45

Arg Thr His Arg Ser Arg Arg Cys Val Ala Pro Val Lys Arg Arg Ile
     50                  55                  60

Gly Cys Ile Lys Ala Val Ala Ala Pro Val Ala Pro Pro Ser Ala Asp
 65                  70                  75                  80

Ser Ala Glu Asp Arg Glu Gln Leu Ala Glu Ser Tyr Gly Phe Arg Gln
                 85                  90                  95
```

Ile Gly Glu Asp Leu Pro Glu Asn Val Thr Leu Lys Asp Ile Met Asp
100 105 110

Thr Leu Pro Lys Glu Val Phe Glu Ile Asp Asp Leu Lys Ala Leu Lys
115 120 125

Ser Val Leu Ile Ser Val Thr Ser Tyr Thr Leu Gly Leu Phe Met Ile
130 135 140

Ala Lys Ser Pro Trp Tyr Leu Leu Pro Leu Ala Trp Ala Trp Thr Gly
145 150 155 160

Thr Ala Ile Thr Gly Phe Phe Val Ile Gly His Asp Cys Ala His Lys
165 170 175

Ser Phe Ser Lys Asn Lys Leu Val Glu Asp Ile Val Gly Thr Leu Ala
180 185 190

Phe Leu Pro Leu Val Tyr Pro Tyr Glu Pro Trp Arg Phe Lys His Asp
195 200 205

Arg His His Ala Lys Thr Asn Met Leu Leu His Asp Thr Ala Trp Gln
210 215 220

Pro Val Pro Pro Glu Glu Phe Glu Ser Ser Pro Val Met Arg Lys Ala
225 230 235 240

Ile Ile Phe Gly Tyr Gly Pro Ile Arg Pro Trp Leu Ser Ile Ala His
245 250 255

Trp Val Asn Trp His Phe Asn Leu Lys Lys Phe Arg Ala Ser Glu Val
260 265 270

Asn Arg Val Lys Ile Ser Leu Ala Cys Val Phe Ala Phe Met Ala Val
275 280 285

Gly Trp Pro Leu Ile Val Tyr Lys Val Gly Ile Leu Gly Trp Val Lys
290 295 300

Phe Trp Leu Met Pro Trp Leu Gly Tyr His Phe Trp Met Ser Thr Phe
305 310 315 320

Thr Met Val His His Thr Ala Pro His Ile Pro Phe Lys Pro Ala Asp
325 330 335

Glu Trp Asn Ala Ala Gln Ala Gln Leu Asn Gly Thr Val His Cys Asp
340 345 350

Tyr Pro Ser Trp Ile Glu Ile Leu Cys His Asp Ile Asn Val His Ile
355 360 365

Pro His His Ile Ser Pro Arg Ile Pro Ser Tyr Asn Leu Arg Ala Ala
370 375 380

His Glu Ser Ile Gln Glu Asn Trp Gly Lys Tyr Thr Asn Leu Ala Thr
385 390 395 400

Trp Asn Trp Arg Leu Met Lys Thr Ile Met Thr Val Cys His Val Tyr
405 410 415

Asp Lys

Claims:

1. A genetically transformed plant which has an elevated linolenic acid content comprising a recombinant, double-stranded DNA molecule comprising

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a structural coding sequence that causes the production of an RNA sequence that encodes a linoleic acid desaturase activity; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

2. The plant of claim 1 in which the linoleic acid desaturase activity is from plants.

3. The plant of claim 1 in which the linoleic acid desaturase activity is from fungi, algae or bacteria.

4. The plant of claim 1 in which the structural coding sequence of (ii) is taken from SEQ. ID NO:1.

5. The plant of claim 1 in which the structural coding sequence of (ii) is taken from SEQ. ID NO:9.

6. The plant of claim 1 in which the structural coding sequence of (ii) is taken from SEQ. ID NO:11.

7. The plant of claim 1 in which the promoter of (i) is an endogenous plant linoleic acid desaturase promoter.

8. A genetically transformed plant which has a reduced linolenic acid content, comprising a recombinant, double-stranded DNA molecule comprising

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a DNA sequence that causes the production of an RNA sequence that is in antisense orientation to at least a portion of a gene that encodes a linoleic acid desaturase activity in said plant; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

9. The plant of claim 8 in which the linoleic acid desaturase enzyme is from plants.

10. The plant of claim 8 in which the linoleic acid desaturase enzyme is from fungi, algae or bacteria.

11. The plant of claim 8 in which the structural coding sequence of (ii) is taken from SEQ. ID NO:1.

12. The plant of claim 8 in which the structural coding sequence of (ii) is taken from SEQ. ID NO:9.

13. The plant of claim 8 in which the structural coding sequence of (ii) is taken from SEQ. 8 ID NO:11.

14. The plant of claim 8 in which the promoter of (i) is an endogenous plant linoleic acid desaturase promoter.

15. A genetically transformed plant which has an improved resistance to low temperatures comprising a recombinant, double-stranded DNA molecule comprising

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a structural coding sequence that causes the production of an RNA sequence that encodes a linoleic acid desaturase activity; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

16. A genetically transformed plant which has an elevated ability to respond to pathogens, comprising a recombinant, double-stranded DNA molecule comprising

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a structural coding sequence that causes the production of an RNA sequence that encodes a linoleic acid desaturase activity; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

17. A seed produced from genetically transformed plant where said seed has an linolenic acid content suitable for use as a source of linolenic acid, said plant comprising a recombinant, double-stranded DNA molecule comprising

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a structural coding sequence that causes the production of an RNA sequence that encodes a linoleic acid desaturase activity; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

18.The seed of claim 17 where said plant is selected from the group consisting of soybean and rapeseed.

19. A genetically transformed plant which has a linolenic acid content of less than about 3%, said plant comprising a recombinant, double-stranded DNA molecule comprising

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a DNA sequence that causes the production of an RNA sequence that is in antisense orientation to at least a portion of a gene that encodes a linoleic acid desaturase activity in said plant; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

20. A genetically transformed plant which has an increased oleic acid content, comprising a recombinant, double-stranded DNA molecule comprising

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a DNA sequence that causes the production of an RNA sequence that is in antisense orientation to at least a portion of a gene that encodes a oleic acid desaturase activity in said plant; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

21. A genetically transformed plant which has an increased oleic acid content, comprising a recombinant, double-stranded DNA molecule comprising

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a DNA sequence that causes the production of an RNA sequence that is in antisense orientation to at least a portion of a gene that encodes a linoleic acid desaturase activity in said plant; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

22. A method of producing a genetically transformed plant which has an elevated linolenic acid content, comprising

(a) inserting into the genome of a plant cell a recombinant, double-stranded DNA molecule comprising:

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a structural coding sequence that causes the production of an RNA sequence that encodes a linoleic acid desaturase activity; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence;

(b) obtaining transformed plant cells; and

(c) regenerating from the transformed plant cells genetically transformed plants which have an elevated linolenic acid content.

23. The method of claim 22 in which the linoleic acid desaturase enzyme is from plants.

24. The method of claim 22 in which the linoleic acid desaturase enzyme is from fungi, algae or bacteria.

25. The method of claim 22 in which the structural coding sequence of (ii) is taken from SEQ. ID NO:1.

26. The method of claim 22 in which the structural coding sequence of (ii) is taken from SEQ. ID NO:9.

27. The method of claim 22 in which the structural coding sequence of (ii) is taken from SEQ. ID NO:11.

28. The plant of claim 22 in which the promoter of (i) is an endogenous plant linoleic acid desaturase promoter.

29. A method of producing a genetically transformed plant which has a reduced linolenic acid content, comprising

(a) inserting into the genome of a plant cell a recombinant, double-stranded DNA molecule comprising:

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a DNA sequence that causes the production of an RNA sequence that is in antisense orientation to at least a portion of a gene that encodes a linoleic acid desaturase activity in said plant; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence

(b) obtaining transformed plant cells; and

(c) regenerating from the transformed plant cells genetically transformed plants which have a reduced linolenic acid content.

30. The method of claim 29 in which the linoleic acid desaturase enzyme is from plants.

31. The method of claim 29 in which the linoleic acid desaturase enzyme is from fungi, algae or bacteria.

32. The method of claim 29 in which the structural coding sequence of (ii) is taken from SEQ. ID NO:1.

33. The method of claim 29 in which the structural coding sequence of (ii) is taken from SEQ. ID NO:9.

34. The method of claim 29 in which the structural coding sequence of (ii) is taken from SEQ. ID NO:11.

35. The plant of claim 29 in which the promoter of (i) is an endogenous plant linoleic acid desaturase promoter.

36. A method of producing a genetically transformed plant which has an increased oleic acid content, comprising

(a) inserting into the genome of a plant cell a recombinant, double-stranded DNA molecule comprising:

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a DNA sequence that causes the production of an RNA sequence that is in antisense orientation to at least a portion of a

gene that encodes a linoleic acid desaturase activity in said plant; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence

(b) obtaining transformed plant cells; and

(c) regenerating from the transformed plant cells genetically transformed plants which have an increased oleic acid content.

37. A recombinant, double-stranded DNA molecule comprising in sequence:

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a structural coding sequence that causes the production of an RNA sequence that encodes a linoleic acid desaturase activity; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

38. A recombinant, double-stranded DNA molecule comprising in sequence:

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a DNA sequence that causes the production of an RNA sequence that is in antisense orientation to at least a portion of a gene that encodes a linoleic acid desaturase activity in said plant; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

39. A plant cell comprising a recombinant, double-stranded DNA molecule comprising in sequence:

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a DNA sequence that causes the production of an RNA sequence that is in antisense orientation to at least a portion of a gene that encodes a linoleic acid desaturase activity in said plant; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence.

40. A method of producing a genetically transformed plant which has an increased oleic acid content, comprising

(a) inserting into the genome of a plant cell a recombinant, double-stranded DNA molecule comprising:

(i) a promoter that functions in plant cells to cause the production of an RNA sequence, said promoter operably linked to;

(ii) a DNA sequence that causes the production of an RNA sequence that is in antisense orientation to at least a portion of a gene that encodes a oleic acid desaturase activity in said plant; and

(iii) a 3' non-translated region that functions in plant cells to promote polyadenylation to the 3' end of said RNA sequence

(b) obtaining transformed plant cells; and

(c) regenerating from the transformed plant cells genetically transformed plants which have an increased oleic acid content.

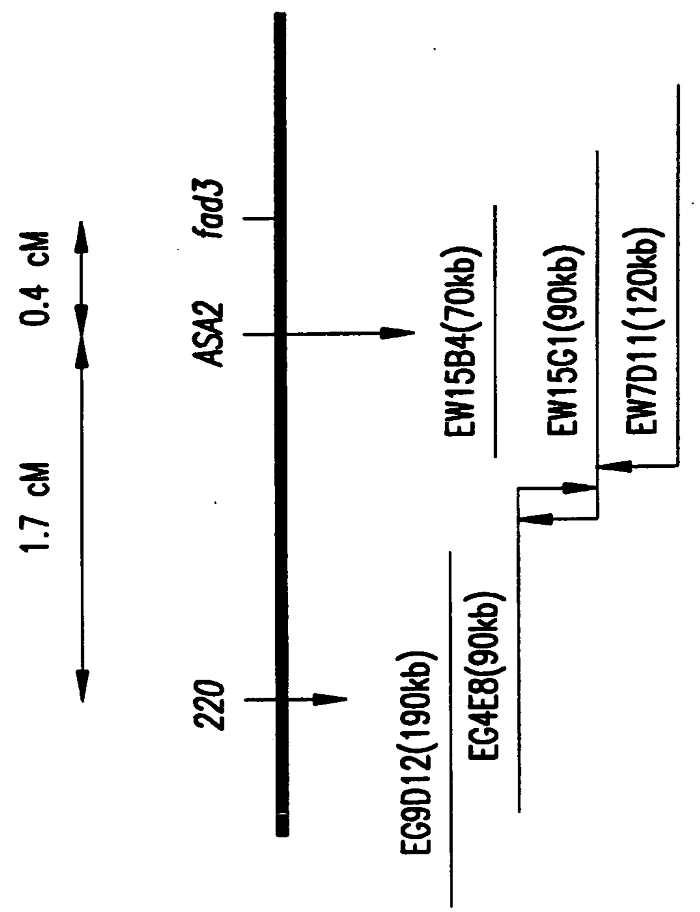

0.4 cM
1.7 cM
fad3
ASA2
220
EW15B4(70kb)
EW15G1(90kb)
EW7D11(120kb)
EG9D12(190kb)
EG4E8(90kb)

FIG.1

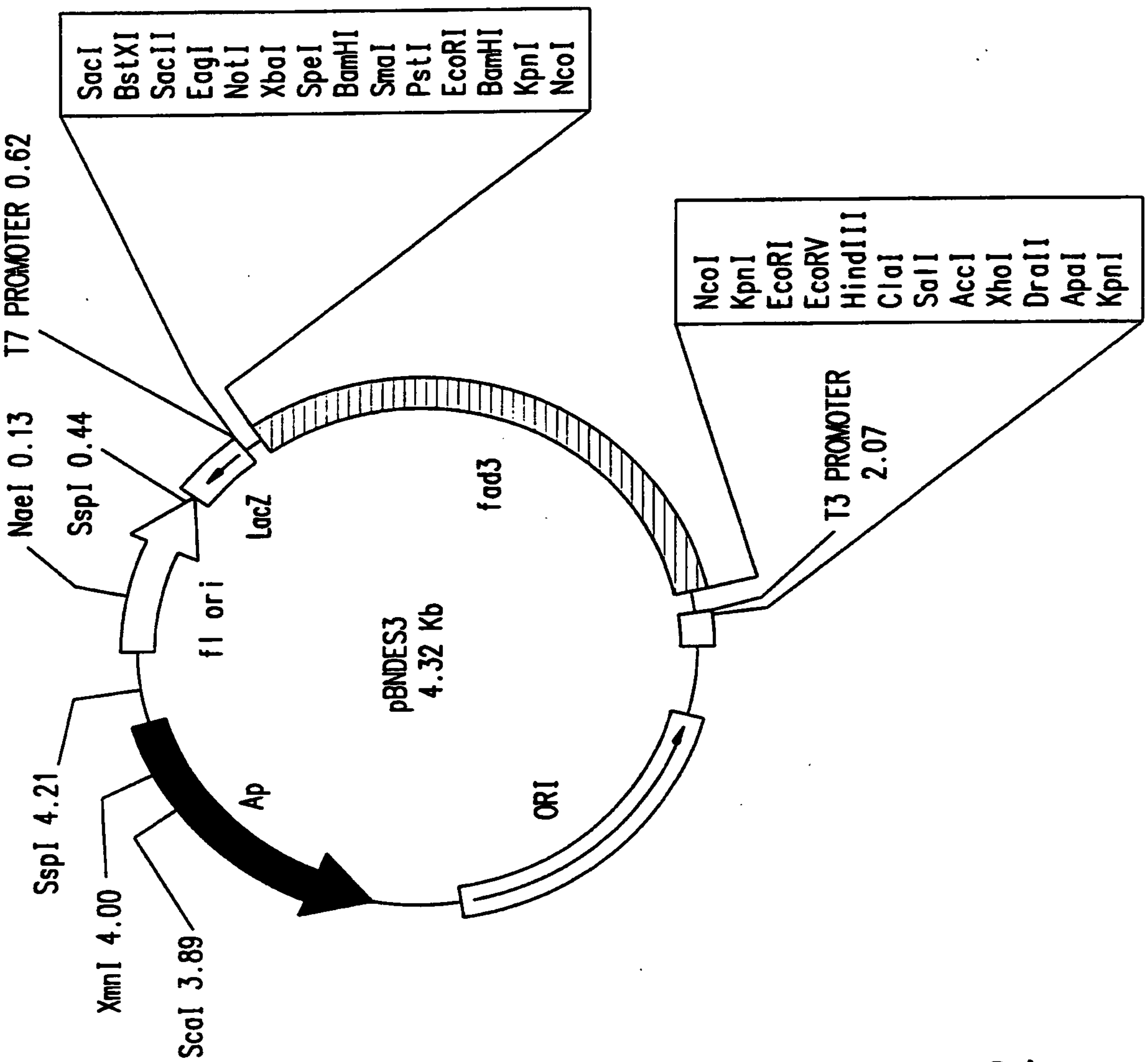

SacI
BstXI
SacII
EagI
NotI
XbaI
SpeI
BamHI
SmaI
PstI
EcoRI
BamHI
KpnI
NcoI
T7 PROMOTER 0.62
NaeI 0.13
SspI 0.44
NcoI
KpnI
EcoRI
EcoRV
HindIII
ClaI
SalI
AccI
XhoI
DraII
ApaI
KpnI
T3 PROMOTER
2.07
LacZ
fad3
fl ori
pBNDES3
4.32 Kb
Ap
ORI
SspI 4.21
XmnI 4.00
ScaI 3.89

FIG.2

AATCCATCAA ACCTTTATTC ACCACATTTC ACTGAAAGGC CACACATCTA GAGAGAGAAA 60

CTTCGTCCAA ATCTCTCTCT CCAGCG ATG GTT GTT GCT ATG GAC CAG CGC AGC 113
Met Val Val Ala Met Asp Gln Arg Ser
1 5

AAT GTT AAC GGA GAT TCC GGT GCC CGG AAG GAA GAA GGG TTT GAT CCA 161
Asn Val Asn Gly Asp Ser Gly Ala Arg Lys Glu Glu Gly Phe Asp Pro
10 15 20 25

AGC GCA CAA CCA CCG TTT AAG ATC GGA GAT ATA AGG GCG GCG ATT CCT 209
Ser Ala Gln Pro Pro Phe Lys Ile Gly Asp Ile Arg Ala Ala Ile Pro
30 35 40

AAG CAT TGC TGG GTG AAG AGT CCT TTG AGA TCT ATG AGC TAC GTC ACC 257
Lys His Cys Trp Val Lys Ser Pro Leu Arg Ser Met Ser Tyr Val Thr
45 50 55

AGA GAC ATT TTC GCC GTC GCG GCT CTG GCC ATG GCC GCC GTG TAT TTT 305
Arg Asp Ile Phe Ala Val Ala Ala Leu Ala Met Ala Ala Val Tyr Phe
60 65 70

GAT AGC TGG TTC CTC TGG CCA CTC TAC TGG GTT GCC CAA GGA ACC CTT 353
Asp Ser Trp Phe Leu Trp Pro Leu Tyr Trp Val Ala Gln Gly Thr Leu
75 80 85

TTC TGG GCC ATC TTC GTT CTT GGC CAC GAC TGT GGA CAT GGG AGT TTC 401
Phe Trp Ala Ile Phe Val Leu Gly His Asp Cys Gly His Gly Ser Phe
90 95 100 105

TCA GAC ATT CCT CTG CTG AAC AGT GTG GTT GGT CAC ATT CTT CAT TCA 449
Ser Asp Ile Pro Leu Leu Asn Ser Val Val Gly His Ile Leu His Ser
110 115 120

TTC ATC CTC GTT CCT TAC CAT GGT TGG AGA ATA AGC CAT CGG ACA CAC 497
Phe Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His Arg Thr His
125 130 135

CAC CAG AAC CAT GGC CAT GTT GAA AAC GAC GAG TCT TGG GTT CCG TTG 545
His Gln Asn His Gly His Val Glu Asn Asp Glu Ser Trp Val Pro Leu
140 145 150

FIG.3a

RECTIFIED SHEET (RULE 91)
ISA/EP

CCA GAA AAG TTG TAC AAG AAC TTG CCC CAT AGT ACT CGG ATG CTC AGA 593
Pro Glu Lys Leu Tyr Lys Asn Leu Pro His Ser Thr Arg Met Leu Arg
155 160 165

TAC ACT GTC CCT CTG CCC ATG CTC GCT TAC CCG ATC TAT CTG TGG TAC 641
Tyr Thr Val Pro Leu Pro Met Leu Ala Tyr Pro Ile Tyr Leu Trp Tyr
170 175 180 185

AGA AGT CCT GGA AAA GAA GGG TCA CAT TTT AAC CCA TAC AGT AGT TTA 689
Arg Ser Pro Gly Lys Glu Gly Ser His Phe Asn Pro Tyr Ser Ser Leu
190 195 200

TTT GCT CCA AGC GAG AGG AAG CTT ATT GCA ACT TCA ACT ACT TGC TGG 737
Phe Ala Pro Ser Glu Arg Lys Leu Ile Ala Thr Ser Thr Thr Cys Trp
205 210 215

TCC ATA ATG TTG GCC ACT CTT GTT TAT CTA TCG TTC CTC GTT GAT CCA 785
Ser Ile Met Leu Ala Thr Leu Val Tyr Leu Ser Phe Leu Val Asp Pro
220 225 230

GTC ACA GTT CTC AAA GTC TAT GGC GTT CCT TAC ATT ATC TTT GTG ATG 833
Val Thr Val Leu Lys Val Tyr Gly Val Pro Tyr Ile Ile Phe Val Met
235 240 245

TGG TTG GAC GCT GTC ACG TAC TTG CAT CAT CAT GGT CAC GAT GAG AAG 881
Trp Leu Asp Ala Val Thr Tyr Leu His His His Gly His Asp Glu Lys
250 255 260 265

TTG CCT TGG TAC AGA GGC AAG GAA TGG AGT TAT TTA CGT GGA GGA TTA 929
Leu Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu
270 275 280

ACA ACT ATT GAT AGA GAT TAC GGA ATC TTC AAC AAC ATC CAT CAC GAC 977
Thr Thr Ile Asp Arg Asp Tyr Gly Ile Phe Asn Asn Ile His His Asp
285 290 295

ATT GGA ACT CAC GTG ATC CAT CAT CTT TTC CCA CAA ATC CCT CAC TAT 1025
Ile Gly Thr His Val Ile His His Leu Phe Pro Gln Ile Pro His Tyr
300 305 310

FIG.3b

RECTIFIED SHEET (RULE 91)
ISA/EP

CAC TTG GTC GAT GCC ACG AGA GCA GCT AAA CAT GTG TTA GGA AGA TAC 1073
His Leu Val Asp Ala Thr Arg Ala Ala Lys His Val Leu Gly Arg Tyr
315 320 325

TAC AGA GAG CCG AAG ACG TCA GGA GCA ATA CCG ATT CAC TTG GTG GAG 1121
Tyr Arg Glu Pro Lys Thr Ser Gly Ala Ile Pro Ile His Leu Val Glu
330 335 340 345

AGT TTG GTC GCA AGT ATT AAA AAA GAT CAT TAC GTC AGT GAC ACT GGT 1169
Ser Leu Val Ala Ser Ile Lys Lys Asp His Tyr Val Ser Asp Thr Gly
350 355 360

GAT ATT GTC TTC TAC GAG ACA GAT CCA GAT CTC TAC GTT TAT GCT TCT 1217
Asp Ile Val Phe Tyr Glu Thr Asp Pro Asp Leu Tyr Val Tyr Ala Ser
365 370 375

GAC AAA TCT AAA ATC AAT TAACTTTTCT TCCTAGCTCT ATTAGGAATA 1265
Asp Lys Ser Lys Ile Asn
380

AACACTCCTT CTCTTTTACT TATTTGTTTC TGCTTTAAGT TTAAAATGTA CTCGTGAAAC 1325

CTTTTTTTTA TTAATGTATT TACGTTAC 1353

FIG.3c

Met Val Val Ala Met Asp Gln Arg Ser Asn Val Asn Gly Asp Ser Gly
1 5 10 15

Ala Arg Lys Glu Glu Gly Phe Asp Pro Ser Ala Gln Pro Pro Phe Lys
20 25 30

Ile Gly Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp Val Lys Ser
35 40 45

Pro Leu Arg Ser Met Ser Tyr Val Thr Arg Asp Ile Phe Ala Val Ala
50 55 60

Ala Leu Ala Met Ala Ala Val Tyr Phe Asp Ser Trp Phe Leu Trp Pro
65 70 75 80

Leu Tyr Trp Val Ala Gln Gly Thr Leu Phe Trp Ala Ile Phe Val Leu
85 90 95

Gly His Asp Cys Gly His Gly Ser Phe Ser Asp Ile Pro Leu Leu Asn
100 105 110

Ser Val Val Gly His Ile Leu His Ser Phe Ile Leu Val Pro Tyr His
115 120 125

Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His Gly His Val
130 135 140

Glu Asn Asp Glu Ser Trp Val Pro Leu Pro Glu Lys Leu Tyr Lys Asn
145 150 155 160

Leu Pro His Ser Thr Arg Met Leu Arg Tyr Thr Val Pro Leu Pro Met
165 170 175

Leu Ala Tyr Pro Ile Tyr Leu Trp Tyr Arg Ser Pro Gly Lys Glu Gly
180 185 190

Ser His Phe Asn Pro Tyr Ser Ser Leu Phe Ala Pro Ser Glu Arg Lys
195 200 205

Leu Ile Ala Thr Ser Thr Thr Cys Trp Ser Ile Met Leu Ala Thr Leu
210 215 220

Val Tyr Leu Ser Phe Leu Val Asp Pro Val Thr Val Leu Lys Val Tyr
225 230 235 240

Gly Val Pro Tyr Ile Ile Phe Val Met Trp Leu Asp Ala Val Thr Tyr
245 250 255

FIG.3d

Leu His His His Gly His Asp Glu Lys Leu Pro Trp Tyr Arg Gly Lys
260 265 270

Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Ile Asp Arg Asp Tyr
275 280 285

Gly Ile Phe Asn Asn Ile His His Asp Ile Gly Thr His Val Ile His
290 295 300

His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Asp Ala Thr Arg
305 310 315 320

Ala Ala Lys His Val Leu Gly Arg Tyr Tyr Arg Glu Pro Lys Thr Ser
325 330 335

Gly Ala Ile Pro Ile His Leu Val Glu Ser Leu Val Ala Ser Ile Lys
340 345 350

Lys Asp His Tyr Val Ser Asp Thr Gly Asp Ile Val Phe Tyr Glu Thr
355 360 365

Asp Pro Asp Leu Tyr Val Tyr Ala Ser Asp Lys Ser Lys Ile Asn
370 375 380

FIG.3e

```
                10        20        30        40        50        60
BND3.AMI    RSNVNGDSGARKEEGFDPSAQPPFKIGDIRAAIPKHCWVKSPLRSMSYVTRDIFAVAALA
                                         .:. ::  ..::.:. :.. .. . :  .. :.:.
DESA.AMI    MTATIPPLTPTVTPSNPDRPIADLKLQDIIKTLPKECFEKKASKAWASVLITLGAIAVGY
                    10        20        30        40        50        60
                70        80        90       100       110       120
BND3.AMI    MAAVYFDSWFLWPLYWVAQGTLFWAIFVLGHDCGHGSFSDIPLLNSVVGHILHSFILVPY
            .. .:. .X.  :. :.  :: . . ::.::::::: ::.. . .:..::::  . .. :.
DESA.AMI    LGIIYL-PWYCLPITWIWTGTALTGAFVVGHDCGHRSFAKKRWVNDLVGHIAFAPLIYPF
                     70        80        90       100       110
               130       140       150       160       170       180
BND3.AMI    HGWRISHRTHHQNHGHVENDESWVPLPEKLYKNLPHSTRMLRYTVPLPH-LAYPIYLWYR
            :.::. :  :: . ...:.:..: : . . ... :  .:..  ... :.  . .:. :
DESA.AMI    HSWRLLHDHHHLHTNKIEVDNAWDFWSVEAFQASPAIVRLFYRAIRGPFWWTGSIFHW--
          120       130       140       150       160       170
                190       200       210       220       230       240
BND3.AMI    SPGKEGSHFNPYSSLFAPSERKLIATSTTCWSIMLATLVYLSFLVDP-V-TVLKVYGVPY
                .  ::.   : ::...:. .  :..   ...:.... ... .. : . V: . .:.
DESA.AMI    ----SLMHFK--LSNFAQRDRNKVKLSIAV-VFLFAAIAFPALIITTGVWGFVKFWLMPW
                  180       190       200       210       220       230
                 250       260       270       280       290       300
BNDS.AMI    IIFVMWLDAVTYLHHHGHDEKLPWYRGKEWSYLRGGL-TTIDRDYGIFNNIH-HDIGTHV
            ... .:... : .::  .. ..    .. .::    . : .:.. ::  . ..  :::..:.
DESA.AMI    LVYHFWMSTFTIVHHTIPEIRF--RPAADWSAAEAQLNGTVHCDYPRWVEVLCHDINVHI
                   240       250       260       270       280
                  310       320       330       340       350       360
BND3.AMI    IHHLFPQIPHYHLVDATRAAKHVLGRYYREPKTSGAIPIHLVESLVASIKKDHYVSDTGD
             :::  .:: :.:  :  . :.  :..
DESA.AMI    PHHLSVAIPSYNLRLAHGSLKENWGPFLYERTFNWQLMQQISGQCHLYDPEHGYRTFGSL
           290       300       310       320       330       340

BND3.AMI    IVF

DESA.AMI    KKV
           350
```

FIG.4

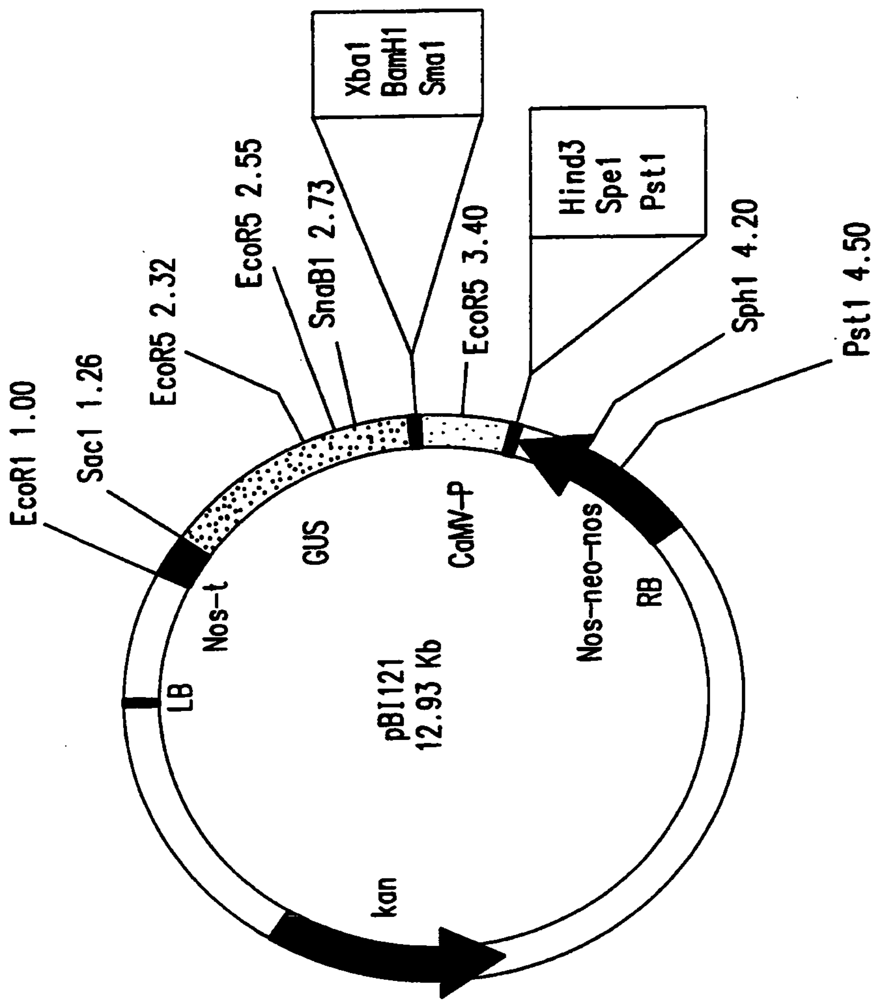

EcoR1 1.00
Sac1 1.26
EcoR5 2.32
EcoR5 2.55
SnaB1 2.73
Xba1
BamH1
Sma1
EcoR5 3.40
Hind3
Spe1
Pst1
Sph1 4.20
Pst1 4.50
GUS
CaMV-P
Nos-neo-nos
Nos-t
RB
LB
pBI121
12.93 Kb
kan

FIG.5

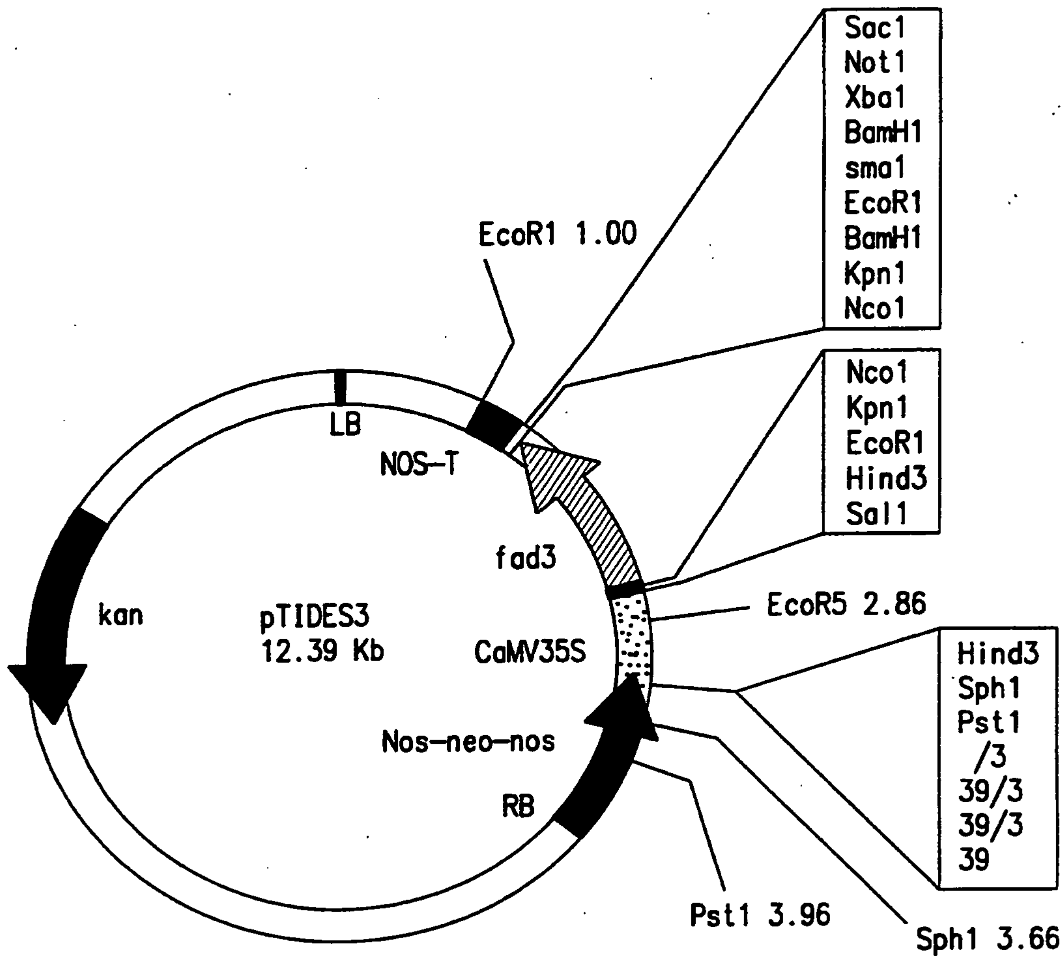

Sac1
Not1
Xba1
BamH1
sma1
EcoR1
BamH1
Kpn1
Nco1
EcoR1 1.00
Nco1
Kpn1
EcoR1
Hind3
Sal1
LB
NOS-T
fad3
kan
pTIDES3
12.39 Kb
CaMV35S
EcoR5 2.86
Hind3
Sph1
Pst1
/3
39/3
39/3
39
Nos-neo-nos
RB
Pst1 3.96
Sph1 3.66

FIG.6

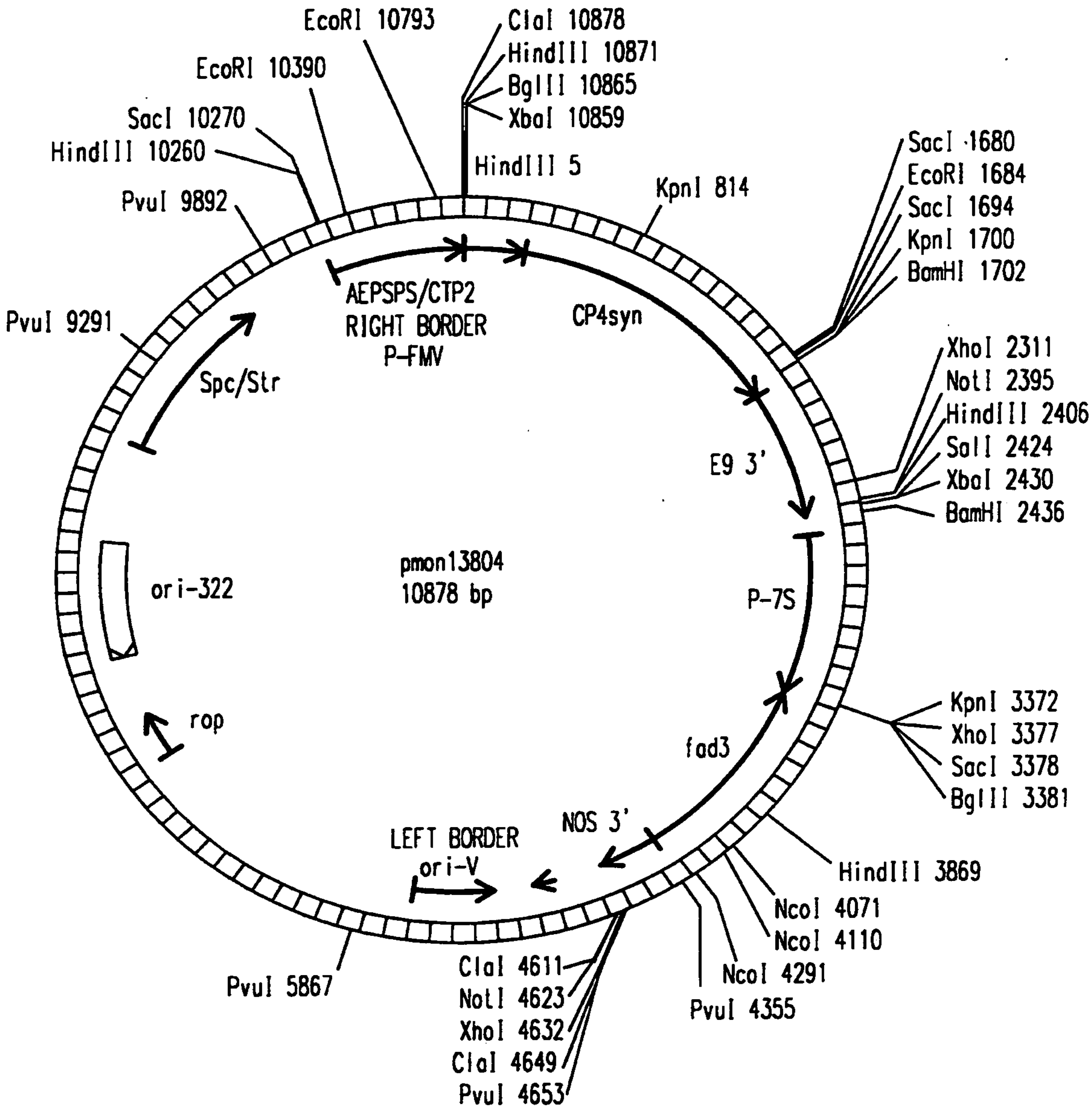


FIG.7

RECTIFIED SHEET (RULE 91)
ISA/EP

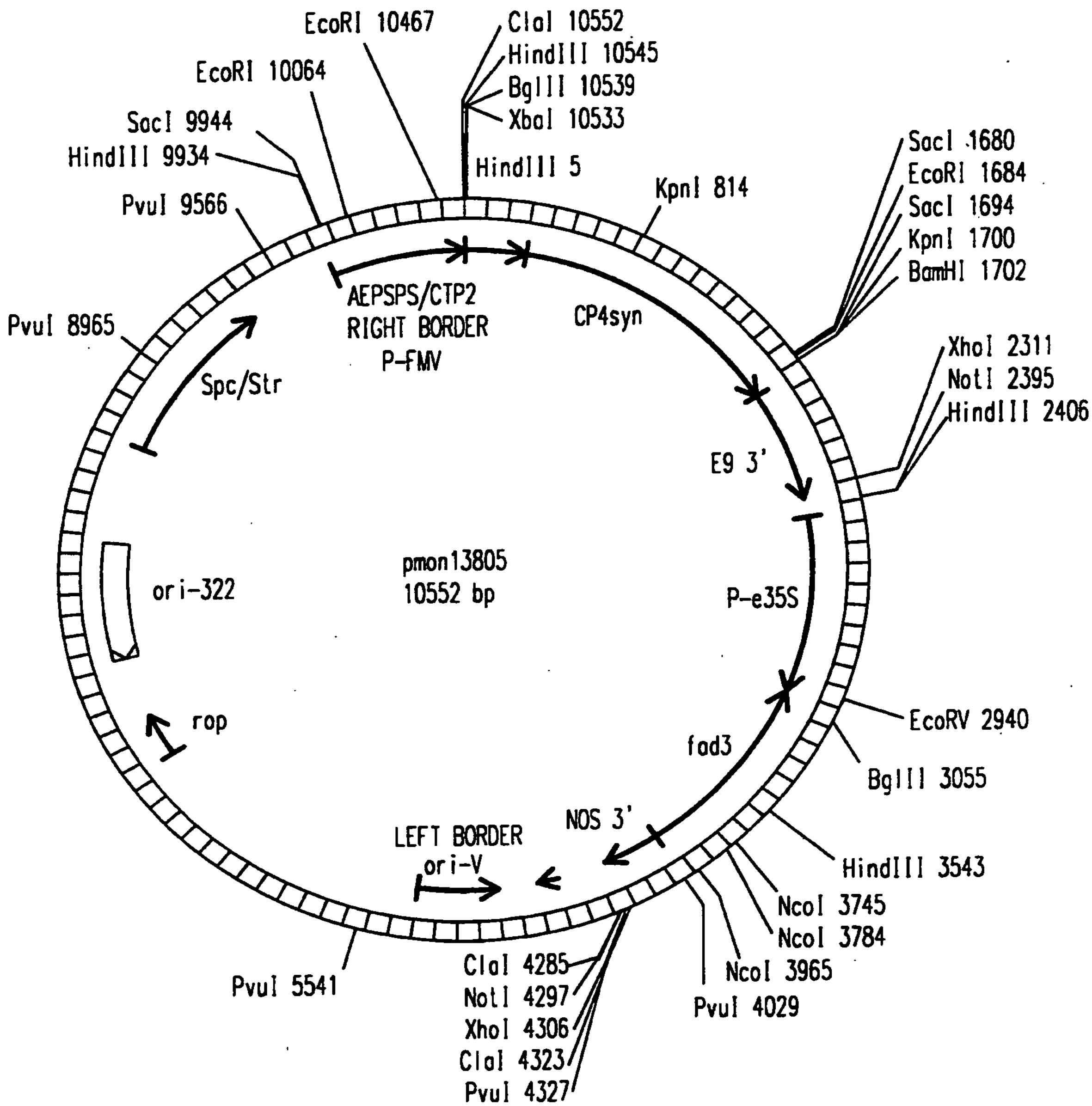


FIG.8

RECTIFIED SHEET (RULE 91)
ISA/EP

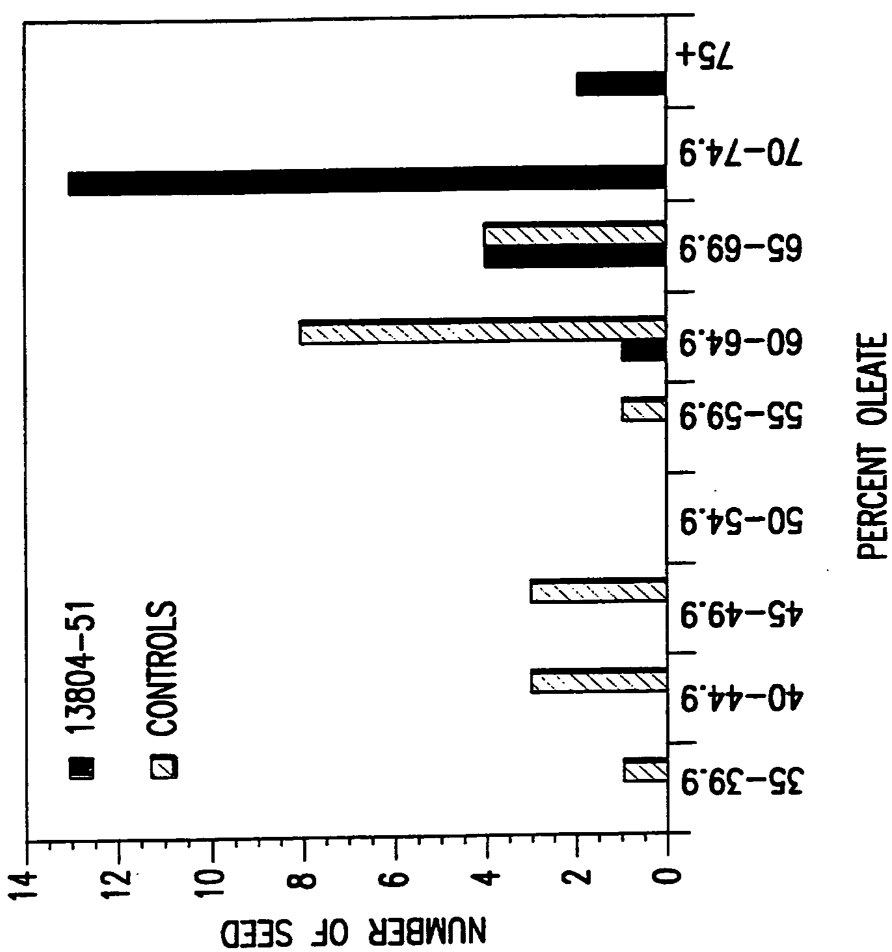

13804–51
CONTROLS
NUMBER OF SEED
14
12
10
8
6
4
2
0
35–39.9
40–44.9
45–49.9
50–54.9
55–59.9
60–64.9
65–69.9
70–74.9
75+
PERCENT OLEATE

FIG.9a

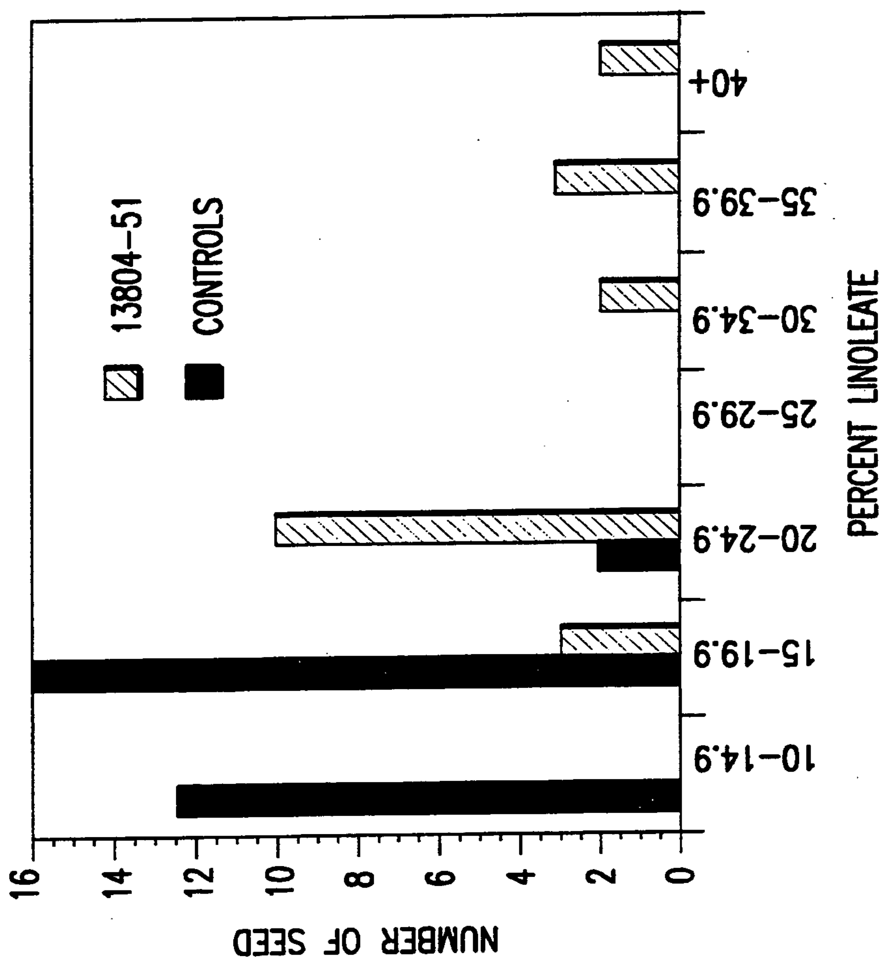

13804-51
CONTROLS
NUMBER OF SEED
16
14
12
10
8
6
4
2
0
10-14.9
15-19.9
20-24.9
25-29.9
30-34.9
35-39.9
40+
PERCENT LINOLEATE

FIG.9b

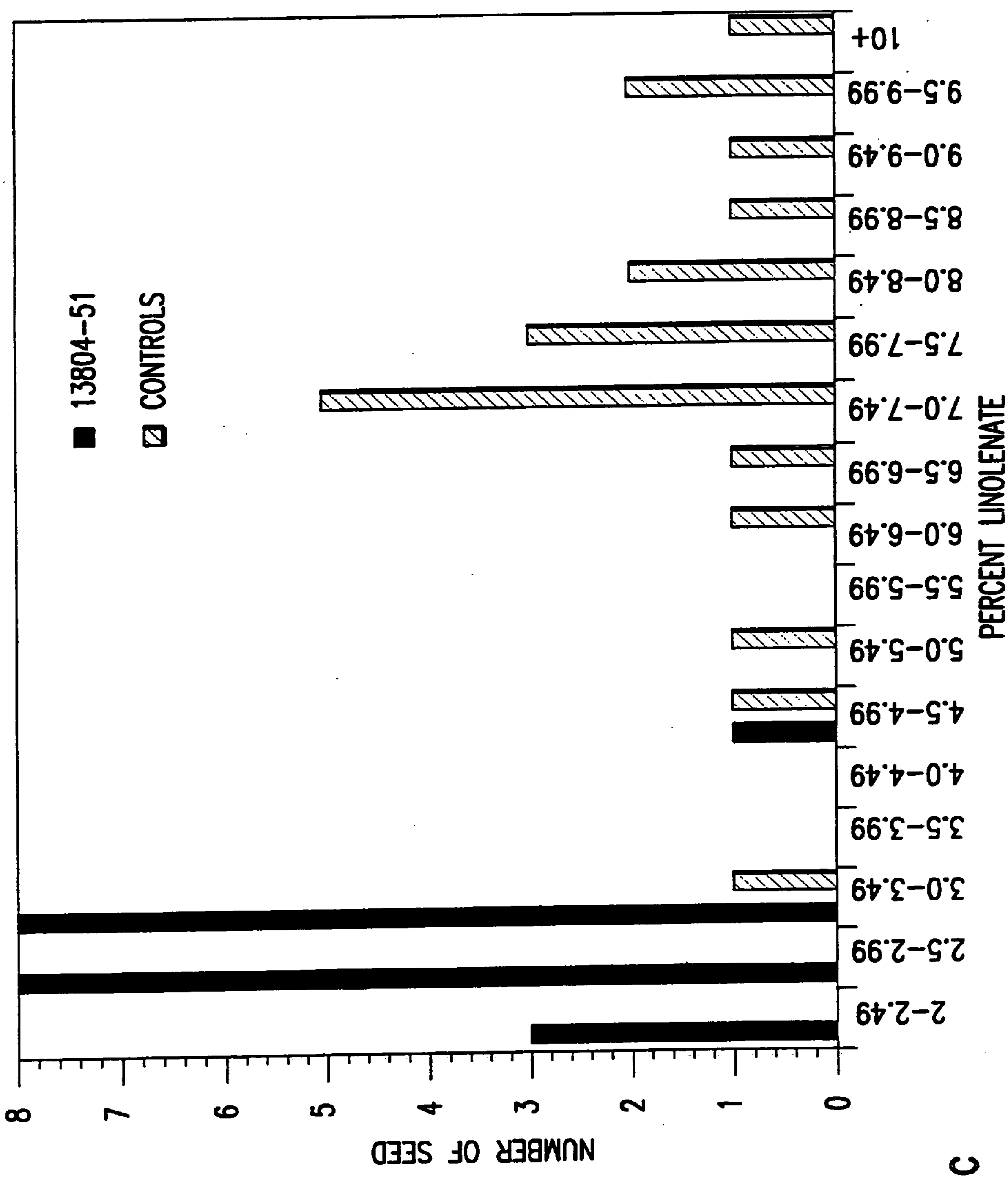

13804-51
CONTROLS
NUMBER OF SEED
0
1
2
3
4
5
6
7
8
PERCENT LINOLENATE
2-2.49
2.5-2.99
3.0-3.49
3.5-3.99
4.0-4.49
4.5-4.99
5.0-5.49
5.5-5.99
6.0-6.49
6.5-6.99
7.0-7.49
7.5-7.99
8.0-8.49
8.5-8.99
9.0-9.49
9.5-9.99
10+

FIG.9c

GGAAAACACA AGTTTCTCTC ACACACATTA TCTCTTTCTC TATTACCACC ACTCATTCAT 60

AACAGAAACC CACCAAAAAA TAAAAAGAGA GACTTTTCAC TCTGGGGAGA GAGCTCAAGT 120

TCTA ATG GCG AAC TTG GTC TTA TCA GAA TGT GGT ATA CGA CCT CTC CCC 169
Met Ala Asn Leu Val Leu Ser Glu Cys Gly Ile Arg Pro Leu Pro
1 5 10 15

AGA ATC TAC ACA ACA CCC AGA TCC AAT TTC CTC TCC AAC AAC AAC AAA 217
Arg Ile Tyr Thr Thr Pro Arg Ser Asn Phe Leu Ser Asn Asn Asn Lys
20 25 30

TTC AGA CCA TCA CTT TCT TCT TCT TCT TAC AAA ACA TCA TCA TCT CCT 265
Phe Arg Pro Ser Leu Ser Ser Ser Ser Tyr Lys Thr Ser Ser Ser Pro
35 40 45

CTG TCT TTT GGT CTG AAT TCA CGA GAT GGG TTC ACG AGG AAT TGG GCG 313
Leu Ser Phe Gly Leu Asn Ser Arg Asp Gly Phe Thr Arg Asn Trp Ala
50 55 60

TTG AAT GTG AGC ACA CCA TTA ACG ACA CCA ATA TTT GAG GAG TCT CCA 361
Leu Asn Val Ser Thr Pro Leu Thr Thr Pro Ile Phe Glu Glu Ser Pro
65 70 75

TTG GAG GAA GAT AAT AAA CAG AGA TTC GAT CCA GGT GCG CCT CCT CCG 409
Leu Glu Glu Asp Asn Lys Gln Arg Phe Asp Pro Gly Ala Pro Pro Pro
80 85 90 95

TTC AAT TTA GCT GAT ATT AGA GCA GCT ATA CCT AAG CAT TGT TGG GTT 457
Phe Asn Leu Ala Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp Val
100 105 110

AAG AAT CCA TGG AAG TCT TTG AGT TAT GTC GTC AGA GAC GTC GCT ATC 505
Lys Asn Pro Trp Lys Ser Leu Ser Tyr Val Val Arg Asp Val Ala Ile
115 120 125

GTC TTT GCA TTG GCT GCT GGA GCT GCT TAC CTC AAC AAT TGG ATT GTT 553
Val Phe Ala Leu Ala Ala Gly Ala Ala Tyr Leu Asn Asn Trp Ile Val
130 135 140

FIG.10a

RECTIFIED SHEET (RULE 91)
ISA/EP

TGG CCT CTC TAT TGG CTC GCT CAA GGA ACC ATG TTT TGG GCT CTC TTT 601
Trp Pro Leu Tyr Trp Leu Ala Gln Gly Thr Met Phe Trp Ala Leu Phe
145 150 155

GTT CTT GGT CAT GAC TGT GGA CAT GGT AGT TTC TCA AAT GAT CCG AAG 649
Val Leu Gly His Asp Cys Gly His Gly Ser Phe Ser Asn Asp Pro Lys
160 165 170 175

TTG AAC AGT GTG GTC GGT CAT CTT CTT CAT TCC TCA ATT CTG GTC CCA 697
Leu Asn Ser Val Val Gly His Leu Leu His Ser Ser Ile Leu Val Pro
180 185 190

TAC CAT GGC TGG AGA ATT AGT CAC AGA ACT CAC CAC CAG AAC CAT GGA 745
Tyr His Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His Gly
195 200 205

CAT GTT GAG AAT GAC GAA TCT TGG CAT CCT ATG TCT GAG AAA ATC TAC 793
His Val Glu Asn Asp Glu Ser Trp His Pro Met Ser Glu Lys Ile Tyr
210 215 220

AAT ACT TTG GAC AAG CCG ACT AGA TTC TTT AGA TTT ACA CTG CCT CTC 841
Asn Thr Leu Asp Lys Pro Thr Arg Phe Phe Arg Phe Thr Leu Pro Leu
225 230 235

GTG ATG CTT GCA TAC CCT TTC TAC TTG TGG GCT CGA AGT CCG GGG AAA 889
Val Met Leu Ala Tyr Pro Phe Tyr Leu Trp Ala Arg Ser Pro Gly Lys
240 245 250 255

AAG GGT TCT CAT TAC CAT CCA GAC AGT GAC TTG TTC CTC CCT AAA GAG 937
Lys Gly Ser His Tyr His Pro Asp Ser Asp Leu Phe Leu Pro Lys Glu
260 265 270

AGA AAG GAT GTC CTC ACT TCT ACT GCT TGT TGG ACT GCA ATG GCT GCT 985
Arg Lys Asp Val Leu Thr Ser Thr Ala Cys Trp Thr Ala Met Ala Ala
275 280 285

CTG CTT GTT TGT CTC AAC TTC ACA ATC GGT CCA ATT CAA ATG CTC AAA 1033
Leu Leu Val Cys Leu Asn Phe Thr Ile Gly Pro Ile Gln Met Leu Lys
290 295 300

FIG.10b

RECTIFIED SHEET (RULE 91)
ISA/EP

CTT TAT GGA ATT CCT TAC TGG ATA AAT GTA ATG TGG TTG GAC TTT GTG 1081
Leu Tyr Gly Ile Pro Tyr Trp Ile Asn Val Met Trp Leu Asp Phe Val
305 310 315

ACT TAC CTG CAT CAC CAT GGT CAT GAA GAT AAG CTT CCT TGG TAC CGT 1129
Thr Tyr Leu His His His Gly His Glu Asp Lys Leu Pro Trp Tyr Arg
320 325 330 335

GGC AAG GAG TGG AGT TAC CTG AGA GGA GGA CTT ACA ACA TTG GAT CGT 1177
Gly Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Leu Asp Arg
340 345 350

GAC TAC GGA TTG ATC AAT AAC ATC CAT CAT GAT ATT GGA ACT CAT GTG 1225
Asp Tyr Gly Leu Ile Asn Asn Ile His His Asp Ile Gly Thr His Val
355 360 365

ATA CAT CAT CTT TTC CCG CAG ATC CCA CAT TAT CAT CTA GTA GAA GCA 1273
Ile His His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Glu Ala
370 375 380

ACA GAA GCA GCT AAA CCA GTA TTA GGG AAG TAT TAC AGG GAG CCT GAT 1321
Thr Glu Ala Ala Lys Pro Val Leu Gly Lys Tyr Tyr Arg Glu Pro Asp
385 390 395

AAG TCT GGA CCG TTG CCA TTA CAT TTA CTG GAA ATT CTA GCG AAA AGT 1369
Lys Ser Gly Pro Leu Pro Leu His Leu Leu Glu Ile Leu Ala Lys Ser
400 405 410 415

ATA AAA GAA GAT CAT TAC GTG AGC GAC GAA GGA GAA GTT GTA TAC TAT 1417
Ile Lys Glu Asp His Tyr Val Ser Asp Glu Gly Glu Val Val Tyr Tyr
420 425 430

AAA GCA GAT CCA AAT CTC TAT GGA GAG GTC AAA GTA AGA GCA GAT TGAAATGAAG 1472
Lys Ala Asp Pro Asn Leu Tyr Gly Glu Val Lys Val Arg Ala Asp
435 440 445

CAGGCTTGAG ATTGAAGTTT TTTCTATTTC AGACCAGCTG ATTTTTTGCT TACTGTATCA 1532

ATTTATTGTG TCACCCACCA GAGAGTTAGT ATCTCTGAAT ACGATCGATC AGATGGAAAC 1592

AACAAATTTG TTTGCGATAC TGAAGCTATA TATACCATAA AAAAAAAAAA AAA 1645

FIG.10c

RECTIFIED SHEET (RULE 91)
ISA/EP

Met Ala Asn Leu Val Leu Ser Glu Cys Gly Ile Arg Pro Leu Pro Arg
1 5 10 15

Ile Tyr Thr Thr Pro Arg Ser Asn Phe Leu Ser Asn Asn Asn Lys Phe
20 25 30

Arg Pro Ser Leu Ser Ser Ser Ser Tyr Lys Thr Ser Ser Ser Pro Leu
35 40 45

Ser Phe Gly Leu Asn Ser Arg Asp Gly Phe Thr Arg Asn Trp Ala Leu
50 55 60

Asn Val Ser Thr Pro Leu Thr Thr Pro Ile Phe Glu Glu Ser Pro Leu
65 70 75 80

Glu Glu Asp Asn Lys Gln Arg Phe Asp Pro Gly Ala Pro Pro Pro Phe
85 90 95

Asn Leu Ala Asp Ile Arg Ala Ala Ile Pro Lys His Cys Trp Val Lys
100 105 110

Asn Pro Trp Lys Ser Leu Ser Tyr Val Val Arg Asp Val Ala Ile Val
115 120 125

Phe Ala Leu Ala Ala Gly Ala Ala Tyr Leu Asn Asn Trp Ile Val Trp
130 135 140

Pro Leu Tyr Trp Leu Ala Gln Gly Thr Met Phe Trp Ala Leu Phe Val
145 150 155 160

Leu Gly His Asp Cys Gly His Gly Ser Phe Ser Asn Asp Pro Lys Leu
165 170 175

Asn Ser Val Val Gly His Leu Leu His Ser Ser Ile Leu Val Pro Tyr
180 185 190

His Gly Trp Arg Ile Ser His Arg Thr His His Gln Asn His Gly His
195 200 205

Val Glu Asn Asp Glu Ser Trp His Pro Met Ser Glu Lys Ile Tyr Asn
210 215 220

FIG.11a

RECTIFIED SHEET (RULE 91)
ISA/EP

Thr Leu Asp Lys Pro Thr Arg Phe Phe Arg Phe Thr Leu Pro Leu Val
225 230 235 240

Met Leu Ala Tyr Pro Phe Tyr Leu Trp Ala Arg Ser Pro Gly Lys Lys
245 250 255

Gly Ser His Tyr His Pro Asp Ser Asp Leu Phe Leu Pro Lys Glu Arg
260 265 270

Lys Asp Val Leu Thr Ser Thr Ala Cys Trp Thr Ala Met Ala Ala Leu
275 280 285

Leu Val Cys Leu Asn Phe Thr Ile Gly Pro Ile Gln Met Leu Lys Leu
290 295 300

Tyr Gly Ile Pro Tyr Trp Ile Asn Val Met Trp Leu Asp Phe Val Thr
305 310 315 320

Tyr Leu His His His Gly His Glu Asp Lys Leu Pro Trp Tyr Arg Gly
325 330 335

Lys Glu Trp Ser Tyr Leu Arg Gly Gly Leu Thr Thr Leu Asp Arg Asp
340 345 350

Tyr Gly Leu Ile Asn Asn Ile His His Asp Ile Gly Thr His Val Ile
355 360 365

His His Leu Phe Pro Gln Ile Pro His Tyr His Leu Val Glu Ala Thr
370 375 380

Glu Ala Ala Lys Pro Val Leu Gly Lys Tyr Tyr Arg Glu Pro Asp Lys
385 390 395 400

Ser Gly Pro Leu Pro Leu His Leu Leu Glu Ile Leu Ala Lys Ser Ile
405 410 415

Lys Glu Asp His Tyr Val Ser Asp Glu Gly Glu Val Val Tyr Tyr Lys
420 425 430

Ala Asp Pro Asn Leu Tyr Gly Glu Val Lys Val Arg Ala Asp
435 440 445

FIG.11b

RECTIFIED SHEET (RULE 91)
ISA/EP

```
AGAGAGTGCA AATAGAACGA CAGAGACTTT TTCCTCTTTT CTTCTTGGGA AGAGGCTCCA      60

ATG GCG AGC TCG GTT TTA TCA GAA TGT GGT TTT AGA CCT CTC CCC AGA       108
Met Ala Ser Ser Val Leu Ser Glu Cys Gly Phe Arg Pro Leu Pro Arg
 1               5                  10                  15

TTC TAC CCT AAA CAC ACA ACC TCT TTT GCC TCT AAC CCT AAA CCC ACT       156
Phe Tyr Pro Lys His Thr Thr Ser Phe Ala Ser Asn Pro Lys Pro Thr
            20                  25                  30

TTC AAA TTC AAT CCA CCA CTT AAA CCT CCT TCT TCT CTT CTC AAT TCC       204
Phe Lys Phe Asn Pro Pro Leu Lys Pro Pro Ser Ser Leu Leu Asn Ser
        35                  40                  45

CGA TAT GGA TTC TAC TCT AAA ACC AGG AAC TGG GCA TTG AAT GTG GCA       252
Arg Tyr Gly Phe Tyr Ser Lys Thr Arg Asn Trp Ala Leu Asn Val Ala
    50                  55                  60

ACA CCT TTA ACA ACT CTT CAG TCT CCA TCC GAG GAA GAC ACG GAG AGA       300
Thr Pro Leu Thr Thr Leu Gln Ser Pro Ser Glu Glu Asp Thr Glu Arg
65                  70                  75                  80

TTC GAC CCA GGT GCG CCT CCT CCC TTC AAT TTG GCG GAT ATA AGA GCA       348
Phe Asp Pro Gly Ala Pro Pro Pro Phe Asn Leu Ala Asp Ile Arg Ala
                85                  90                  95

GCC ATA CCT AAG CAT TGT TGG GTT AAG AAT CCA TGG ATG TCT ATG AGT       396
Ala Ile Pro Lys His Cys Trp Val Lys Asn Pro Trp Met Ser Met Ser
            100                 105                 110

TAT GTT GTC AGA GAT GTT GCT ATC GTC TTT GGA TTG GCT GCT GTT GCT       444
Tyr Val Val Arg Asp Val Ala Ile Val Phe Gly Leu Ala Ala Val Ala
        115                 120                 125

GCT TAC TTC AAC AAT TGG CTT CTC TGG CCT CTC TAC TGG TTC GCT CAA       492
Ala Tyr Phe Asn Asn Trp Leu Leu Trp Pro Leu Tyr Trp Phe Ala Gln
    130                 135                 140

GGA ACC ATG TTC TGG GCT CTC TTT GTC CTT GGC CAT GAC TGC GGA CAT       540
Gly Thr Met Phe Trp Ala Leu Phe Val Leu Gly His Asp Cys Gly His
145                 150                 155                 160
```

FIG.12a

GGT AGC TTC TCG AAT GAT CCG AGG CTG AAC AGT GTG GCT GGT CAT CTT 588
Gly Ser Phe Ser Asn Asp Pro Arg Leu Asn Ser Val Ala Gly His Leu
165 170 175

CTT CAT TCC TCA ATT CTG GTC CCT TAC CAT GGC TGG AGG ATT AGC CAC 636
Leu His Ser Ser Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His
180 185 190

AGA ACT CAC CAC CAG AAC CAT GGT CAT GTC GAG AAT GAC GAA TCA TGG 684
Arg Thr His His Gln Asn His Gly His Val Glu Asn Asp Glu Ser Trp
195 200 205

CAT CCT TTG CCT GAA AGC ATC TAC AAG AAT TTG GAA AAG ACG ACT CAA 732
His Pro Leu Pro Glu Ser Ile Tyr Lys Asn Leu Glu Lys Thr Thr Gln
210 215 220

ATG TTT AGG TTT ACA CTG CCT TTT CCA ATG CTC GCA TAC CCT TTC TAC 780
Met Phe Arg Phe Thr Leu Pro Phe Pro Met Leu Ala Tyr Pro Phe Tyr
225 230 235 240

TTG TGG AAC AGA AGT CCA GGG AAA CAA GGT TCT CAT TAT CAT CCG GAC 828
Leu Trp Asn Arg Ser Pro Gly Lys Gln Gly Ser His Tyr His Pro Asp
245 250 255

AGT GAC TTG TTT CTT CCA AAA GAG AAG AAA GAT GTT CTG ACA TCA ACT 876
Ser Asp Leu Phe Leu Pro Lys Glu Lys Lys Asp Val Leu Thr Ser Thr
260 265 270

GCC TGT TGG ACT GCA ATG GCT GCT TTG CTT GTT TGT CTC AAC TTT GTC 924
Ala Cys Trp Thr Ala Met Ala Ala Leu Leu Val Cys Leu Asn Phe Val
275 280 285

ATG GGT CCA ATC CAG ATG CTC AAA CTA TAT GGC ATC CCT TAT TGG ATA 972
Met Gly Pro Ile Gln Met Leu Lys Leu Tyr Gly Ile Pro Tyr Trp Ile
290 295 300

TTT GTA ATG TGG TTG GAC TTC GTC ACT TAC TTG CAC CAC CAT GGA CAT 1020
Phe Val Met Trp Leu Asp Phe Val Thr Tyr Leu His His His Gly His
305 310 315 320

FIG.12b

RECTIFIED SHEET (RULE 91)
ISA/EP

```
GAA GAC AAG CTC CCT TGG TAT CGT GGA AAG GAA TGG AGT TAC CTG AGA    1068
Glu Asp Lys Leu Pro Trp Tyr Arg Gly Lys Glu Trp Ser Tyr Leu Arg
                325                 330                 335

GGA GGG CTC ACA ACA TTA GAT CGT GAC TAC GGA TGG ATC AAT AAC ATC    1116
Gly Gly Leu Thr Thr Leu Asp Arg Asp Tyr Gly Trp Ile Asn Asn Ile
            340                 345                 350

CAC CAC GAT ATT GGA ACT CAT GTG ATA CAT CAT CTT TTC CCG CAG ATC    1164
His His Asp Ile Gly Thr His Val Ile His His Leu Phe Pro Gln Ile
        355                 360                 365

CCA CAT TAT CAT CTA GTA GAA GCA ACA GAA GCA GCT AAA CCA GTA CTA    1212
Pro His Tyr His Leu Val Glu Ala Thr Glu Ala Ala Lys Pro Val Leu
    370                 375                 380

GGA AAG TAC TAC AGA GAA CCG AAA AAC TCT GGA CCT CTG CCA CTT CAC    1260
Gly Lys Tyr Tyr Arg Glu Pro Lys Asn Ser Gly Pro Leu Pro Leu His
385                 390                 395                 400

TTA CTG GGA AGC CTC ATA AAG AGT ATG AAA CAA GAC CAT TTC GTA AGC    1308
Leu Leu Gly Ser Leu Ile Lys Ser Met Lys Gln Asp His Phe Val Ser
                405                 410                 415

GAT ACA GGA GAT GTC GTG TAC TAT GAG GCA GAT CCA AAA CTC AAT GGA    1356
Asp Thr Gly Asp Val Val Tyr Tyr Glu Ala Asp Pro Lys Leu Asn Gly
            420                 425                 430

CAA AGA ACA TGAGGACATA CTGCAGTGAA CCAGGCAGAC AAGTTACATA            1405
Gln Arg Thr
        435

AATTCATCTT GGCCCATTCA TTATGTTCTT TTTGTTTTGG TGTAAAGCCT TTTCGAGATT  1465

AAAAAAGCAT TAATTTGTAG AAACCTGTGG TAAAACTCTC GATCAAATGA AATAAGATAT  1525
```

FIG.12c

RECTIFIED SHEET (RULE 91)
ISA/EP

Met Ala Ser Ser Val Leu Ser Glu Cys Gly Phe Arg Pro Leu Pro Arg
1 5 10 15

Phe Tyr Pro Lys His Thr Thr Ser Phe Ala Ser Asn Pro Lys Pro Thr
20 25 30

Phe Lys Phe Asn Pro Pro Leu Lys Pro Pro Ser Ser Leu Leu Asn Ser
35 40 45

Arg Tyr Gly Phe Tyr Ser Lys Thr Arg Asn Trp Ala Leu Asn Val Ala
50 55 60

Thr Pro Leu Thr Thr Leu Gln Ser Pro Ser Glu Glu Asp Thr Glu Arg
65 70 75 80

Phe Asp Pro Gly Ala Pro Pro Pro Phe Asn Leu Ala Asp Ile Arg Ala
85 90 95

Ala Ile Pro Lys His Cys Trp Val Lys Asn Pro Trp Met Ser Met Ser
100 105 110

Tyr Val Val Arg Asp Val Ala Ile Val Phe Gly Leu Ala Ala Val Ala
115 120 125

Ala Tyr Phe Asn Asn Trp Leu Leu Trp Pro Leu Tyr Trp Phe Ala Gln
130 135 140

Gly Thr Met Phe Trp Ala Leu Phe Val Leu Gly His Asp Cys Gly His
145 150 155 160

Gly Ser Phe Ser Asn Asp Pro Arg Leu Asn Ser Val Ala Gly His Leu
165 170 175

Leu His Ser Ser Ile Leu Val Pro Tyr His Gly Trp Arg Ile Ser His
180 185 190

Arg Thr His His Gln Asn His Gly His Val Glu Asn Asp Glu Ser Trp
195 200 205

His Pro Leu Pro Glu Ser Ile Tyr Lys Asn Leu Glu Lys Thr Thr Gln
210 215 220

FIG.13a

RECTIFIED SHEET (RULE 91)
ISA/EP

Met	Phe	Arg	Phe	Thr	Leu	Pro	Phe	Pro	Met	Leu	Ala	Tyr	Pro	Phe	Tyr
225					230					235					240
Leu	Trp	Asn	Arg	Ser	Pro	Gly	Lys	Gln	Gly	Ser	His	Tyr	His	Pro	Asp
				245					250					255	
Ser	Asp	Leu	Phe	Leu	Pro	Lys	Glu	Lys	Lys	Asp	Val	Leu	Thr	Ser	Thr
			260					265					270		
Ala	Cys	Trp	Thr	Ala	Met	Ala	Ala	Leu	Leu	Val	Cys	Leu	Asn	Phe	Val
		275					280					285			
Met	Gly	Pro	Ile	Gln	Met	Leu	Lys	Leu	Tyr	Gly	Ile	Pro	Tyr	Trp	Ile
	290					295					300				
Phe	Val	Met	Trp	Leu	Asp	Phe	Val	Thr	Tyr	Leu	His	His	His	Gly	His
305					310					315					320
Glu	Asp	Lys	Leu	Pro	Trp	Tyr	Arg	Gly	Lys	Glu	Trp	Ser	Tyr	Leu	Arg
				325					330					335	
Gly	Gly	Leu	Thr	Thr	Leu	Asp	Arg	Asp	Tyr	Gly	Trp	Ile	Asn	Asn	Ile
			340					345					350		
His	His	Asp	Ile	Gly	Thr	His	Val	Ile	His	His	Leu	Phe	Pro	Gln	Ile
		355					360					365			
Pro	His	Tyr	His	Leu	Val	Glu	Ala	Thr	Glu	Ala	Ala	Lys	Pro	Val	Leu
	370					375					380				
Gly	Lys	Tyr	Tyr	Arg	Glu	Pro	Lys	Asn	Ser	Gly	Pro	Leu	Pro	Leu	His
385					390					395					400
Leu	Leu	Gly	Ser	Leu	Ile	Lys	Ser	Met	Lys	Gln	Asp	His	Phe	Val	Ser
				405					410					415	
Asp	Thr	Gly	Asp	Val	Val	Tyr	Tyr	Glu	Ala	Asp	Pro	Lys	Leu	Asn	Gly
			420					425					430		
Gln	Arg	Thr													
		435													

FIG.13b

RECTIFIED SHEET (RULE 91)
ISA/EP

INTERNATIONAL SEARCH REPORT

Inte. ...al Application No
PCT/US 94/01321

A. CLASSIFICATION OF SUBJECT MATTER
IPC 5 C12N15/82 C12N15/53 C12N15/11 C12N5/10 A01H5/00
C11B1/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 5 C12N A01H C11B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SCIENCE vol. 258, 20 November 1992, LANCASTER, PA US pages 1353 - 1355 ARONDEL, V., ET AL. 'Map-based cloning of a gene controlling omega-3 fatty acid desaturation in Arabidopsis'	37
Y	see the whole document	1,2,4,8, 9,11,17, 18,22, 23,25, 29,30, 32,38

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
1 June 1994	14 -06- 1994

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Maddox, A

International Application No
PCT/US 94/01321

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	US,A,5 057 419 (MARTIN) 15 October 1991 see column 6, line 40 - column 6, line 66 see column 9, line 25 - column 10, line 58	20,40 1,2,4,8, 9,11,17, 18,22, 23,25, 29,30, 32,38
Y	WO,A,91 13972 (CALGENE) 19 September 1991 see the whole document	1,2,4,8, 9,11,17, 18,22, 23,25, 29,30, 32,38
P,X	JOURNAL OF BIOLOGICAL CHEMISTRY vol. 268, no. 32 , 15 November 1993 , BALTIMORE, MD US pages 24099 - 24105 IBA, K., ET AL. 'A gene encoding a chloroplast omega-3 fatty acid desaturase complements alterations in fatty acid desaturation and chloroplast copy numbers of the fad7 mutant of Arabidopsis thaliana' see the whole document	1,2,5, 22,23, 26,37
P,X	PLANT PHYSIOLOGY. vol. 103 , October 1993 , ROCKVILLE, MD, USA. pages 467 - 476 YADAV, N.S., ET AL. 'Cloning of higher plant omega-3- fatty acid desaturases' see the whole document	1,2,17, 22,23,37
P,X	WO,A,93 11245 (DU PONT) 10 June 1993 see the whole document	1,2,8,9, 17,22, 23,29, 30,37,38
P,X	WO,A,93 06712 (RHONE-POULENC AGROCHIMIE) 15 April 1993	1,2,22, 23,37
A	PLANT PHYSIOLOGY. vol. 100 , 1992 , ROCKVILLE, MD, USA. pages 894 - 901 POLASHOCK, J.J., ET AL. 'Expression of the yeast delta-9 fatty acid desaturase in Nicotina tabacum' see the whole document	1,22

-/--

1

Inte. nal Application No
PCT/US 94/01321

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ANN. REV. PLANT PHYSIOL. PLANT MOL. BIOL. vol. 42 , 1991 pages 467 - 506 BROWSE, J., ET AL. 'Glycerolipid synthesis: Biochemistry and regulation' see the whole document ---	1-40
A	UCLA SYMP. MOL. CELL BIOL, NEW SER. vol. 129 , 1990 pages 301 - 309 BROWSE, J., ET AL. 'Strtegies for modifying plant lipid composition' see page 306 ---	1,22
A	NL,A,9 002 130 (STICHTING TECHNISCHE WETENSCHAPPEN UTRECHT) 16 April 1992 see the whole document -----	1-40

1

INTERNATIONAL SEARCH REPORT

Inte... ...al Application No
PCT/US 94/01321

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
US-A-5057419	15-10-91	NONE		
WO-A-9113972	19-09-91	EP-A-	0472722	04-03-92
WO-A-9311245	10-06-93	AU-A-	3228893	28-06-93
WO-A-9306712	15-04-93	AU-A-	2881292	03-05-93
NL-A-9002130	16-04-92	NONE		

CORRECTED VERSION*

55
52

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/82, 15/29, 15/53, A01H 5/00, 5/10

A1

(11) International Publication Number: WO 98/45461

(43) International Publication Date: 15 October 1998 (15.10.98)

(21) International Application Number: PCT/US98/07179

(22) International Filing Date: 9 April 1998 (09.04.98)

(30) Priority Data:
08/831,575 9 April 1997 (09.04.97) US

(71) Applicant *(for all designated States except US)*: RHONE-POULENC AGRO [FR/FR]; Dépt. Propriété Industrielle, 14-20, rue Pierre Baizet, F-69009 Lyon (FR).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: THOMAS, Terry, L. [US/US]; 2804 Cloister Drive, College Station, TX 77845 (US). LI, Zhongsen [CN/US]; Apartment Z-1-H, 1 Hensel, College Station, TX 77840 (US).

(74) Agents: DiGIGLIO, Frank, S. et al.; Scully, Scott, Murphy & Presser, 400 Garden City Plaza, Garden City, NY 11530 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: AN OLEOSIN 5' REGULATORY REGION FOR THE MODIFICATION OF PLANT SEED LIPID COMPOSITION

(57) Abstract

The present invention is directed to 5' regulatory regions of an *Arabidopsis* oleosin gene. The 5' regulatory regions, when operably linked to either the coding sequence of a heterologous gene or a sequence complementary to a native plant gene, direct expression of the coding sequence or complementary sequence in a plant seed. The regulatory regions are useful in expression cassettes and expression vectors for the transformation of plants. Also provided are methods of modulating the levels of a heterologous gene such as a fatty acid synthesis or lipid metabolism gene by transforming a plant with the subject expression cassettes and expression vectors.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

AN OLEOSIN 5' REGULATORY REGION FOR THE MODIFICATION OF PLANT SEED LIPID COMPOSITION

BACKGROUND OF THE INVENTION

Seed oil content has traditionally been modified by plant breeding. The use of recombinant DNA technology to alter seed oil composition can accelerate this process and in some cases alter seed oils in a way that cannot be accomplished by breeding alone. The oil composition of *Brassica* has been significantly altered by modifying the expression of a number of lipid metabolism genes. Such manipulations of seed oil composition have focused on altering the proportion of endogenous component fatty acids. For example, antisense repression of the Δ12-desaturase gene in transgenic rapeseed has resulted in an increase in oleic acid of up to 83%. Topfer et al. 1995 *Science 268*:681-686.

There have been some successful attempts at modifying the composition of seed oil in transgenic plants by introducing new genes that allow the production of a fatty acid that the host plants were not previously capable of synthesizing. Van de Loo, et al. (1995 *Proc. Natl. Acad. Sci USA 92*:6743-6747) have been able to introduce a Δ12-hydroxylase gene into transgenic tobacco, resulting in the introduction of a novel fatty acid, ricinoleic acid, into its seed oil. The reported accumulation was modest from plants carrying constructs in which transcription of the hydroxylase gene was under the control of the cauliflower mosaic virus (CaMV) 35S promoter.

Similarly, tobacco plants have been engineered to produce low levels of petroselinic acid by expression of an acyl-ACP desaturase from coriander (Cahoon et al. 1992 *Proc. Natl. Acad. Sci USA 89*:11184-11188).

The long chain fatty acids (C18 and larger), have significant economic value both as nutritionally and medically important foods and as industrial commodities (Ohlrogge,J.B. 1994 *Plant Physiol. 104*:821-826). Linoleic (18:2 Δ9,12) and α-linolenic acid (18:3 Δ9,12,15) are essential fatty acids found in many seed oils. The levels of these fatty-acids have been manipulated in oil seed crops through breeding and biotechnology (Ohlrogge, et al. *1991 Biochim. Biophys. Acta 1082*:1-26; Topfer et al. 1995 *Science 268*:681-686). Additionally, the production of novel fatty acids in seed oils can be of considerable use in both human health and industrial applications.

Consumption of plant oils rich in γ-linolenic acid (GLA) (18:3 Δ6,9,12) is thought to alleviate hypercholesterolemia and other related clinical disorders which correlate with susceptibility to coronary heart disease (Brenner R.R. 1976 *Adv. Exp. Med. Biol. 83*:85-101). The therapeutic benefits of dietary GLA may result from its role as a precursor to prostaglandin synthesis (Weete, J.D. 1980 in *Lipid Biochemistry of Fungi and Other Organisms*, eds. Plenum Press, New York, pp. 59-62). Linoleic acid(18:2) (LA) is transformed into gamma linolenic acid (18:3) (GLA) by the enzyme Δ6-desaturase.

Few seed oils contain GLA despite high contents of the precursor linoleic acid. This is due to the absence of Δ6-desaturase activity in most plants. For example, only borage (*Borago officinalis*), evening primrose (*Oenothera biennis*), and currants (*Ribes nigrum*) produce appreciable amounts of linolenic acid. Of these three species, only *Oenothera* and Borage are cultivated as a commercial source for GLA. It would be beneficial if agronomic seed oils could be engineered to produce GLA in significant quantities by introducing a heterologous Δ6-desaturase gene. It would also be beneficial if other expression products associated with fatty acid synthesis and lipid metabolism could be produced in plants at high enough levels so that commercial production of a particular expression product becomes feasible.

As disclosed in U.S. Patent No. 5,552,306, a cyanobacterial Δ^6-desaturase gene has been recently isolated. Expression of this cyanobacterial gene in transgenic tobacco resulted in significant but low level GLA accumulation. (Reddy et al. 1996 *Nature Biotech. 14*:639-642). Applicant's copending U.S. Application Serial No. 08,366,779, discloses a Δ6-desaturase gene isolated from the plant *Borago officinalis* and its expression in tobacco under the control of the CaMV 35S promoter. Such expression resulted in significant but low level GLA and octadecatetraenoic acid (ODTA or OTA) accumulation in seeds. Thus, a need exists for a promoter which

functions in plants and which consistently directs high level expression of lipid metabolism genes in transgenic plant seeds.

Oleosins are abundant seed proteins associated with the phospholipid monolayer membrane of oil bodies. The first oleosin gene, L3, was cloned from maize by selecting clones whose in vitro translated products were recognized by an anti-L3 antibody (Vance et al. 1987 *J. Biol. Chem. 262:11275-11279*). Subsequently, different isoforms of oleosin genes from such different species as *Brassica*, soybean, carrot, pine, and *Arabidopsis* have been cloned (Huang, A.H.C., 1992, *Ann. Reviews Plant Phys. and Plant Mol. Biol. 43*:177-200; Kirik et al., 1996 *Plant Mol. Biol. 31*:413-417; Van Rooijen et al., 1992 *Plant Mol. Biol. 18*:1177-1179; Zou et al., *Plant Mol. Biol. 31*:429-433. Oleosin protein sequences predicted from these genes are highly conserved, especially for the central hydrophobic domain. All of these oleosins have the characteristic feature of three distinctive domains. An amphipathic domain of 40-60 amino acids is present at the N-terminus; a totally hydrophobic domain of 68-74 amino acids is located at the center; and an amphipathic α-helical domain of 33-40 amino acids is situated at the C-terminus (Huang, A.H.C. 1992).

The present invention provides 5' regulatory sequences from an oleosin gene which direct high level expression of lipid metabolism genes in transgenic plants. In accordance with the present invention,

chimeric constructs comprising an oleosin 5' regulatory region operably linked to coding sequence for a lipid metabolism gene such as a Δ6-desaturase gene are provided. Transgenic plants comprising the subject chimeric constructs produce levels of GLA approaching the level found in those few plant species which naturally produce GLA such as evening primrose (*Oenothera biennis*).

SUMMARY OF THE INVENTION

The present invention is directed to 5' regulatory regions of an *Arabidopsis* oleosin gene. The 5' regulatory regions, when operably linked to either the coding sequence of a heterologous gene or sequence complementary to a native plant gene, direct expression of the heterologous gene or complementary sequence in a plant seed.

The present invention thus provides expression cassettes and expression vectors comprising an oleosin 5' regulatory region operably linked to a heterologous gene or a sequence complementary to a native plant gene.

Plant transformation vectors comprising the expression cassettes and expression vectors are also provided as are plant cells transformed by these vectors, and plants and their progeny containing the vectors.

In one embodiment of the invention, the heterologous gene or complementary gene sequence is a fatty acid synthesis gene or a lipid metabolism gene.

In another aspect of the present invention, a method is provided for producing a plant with increased levels of a product of a fatty acid synthesis or lipid metabolism gene.

In particular, there is provided a method for producing a plant with increased levels of a fatty acid synthesis or lipid metabolism gene by transforming a plant with the subject expression cassettes and expression vectors which comprise an oleosin 5' regulatory region and a coding sequence for a fatty acid synthesis or lipid metabolism gene.

In another aspect of the present invention, there is provided a method for cosuppressing a native fatty acid synthesis or lipid metabolism gene by transforming a plant with the subject expression cassettes and expression vectors which comprise an oleosin 5' regulatory region and a coding sequence for a fatty acid synthesis or lipid metabolism gene.

A further aspect of this invention provides a method of decreasing production of a native plant gene such as a fatty acid synthesis gene or a lipid metabolism gene by transforming a plant with an expression vector comprising a oleosin 5' regulatory region operably linked to a nucleic acid sequence complementary to a native plant gene.

Also provided are methods of modulating the levels of a heterologous gene such as a fatty acid synthesis or lipid metabolism gene by transforming a plant with the subject expression cassettes and expression vectors.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 depicts the nucleotide and corresponding amino acid sequence of the borage Δ6-desaturase gene (SEQ ID NO:1). The cytochrome b5 heme-binding motif is boxed and the putative metal binding, histidine rich motifs (HRMs) are underlined. The motifs recognized by the primers (PCR analysis) are underlined with dotted lines, i.e. tgg aaa tgg aac cat aa; and gag cat cat ttg ttt cc.

Fig. 2 is a dendrogram showing similarity of the borage Δ6-desaturase to other membrane-bound desaturases. The amino acid sequence of the borage Δ6-desaturase was compared to other known desaturases using Gene Works (IntelliGenetics). Numerical values correlate to relative phylogenetic distances between subgroups compared.

Fig. 3A provides a gas liquid chromatography profile of the fatty acid methyl esters (FAMES) derived from leaf tissue of a wild type tobacco 'Xanthi'.

Fig. 3B provides a gas liquid chromatography profile of the FAMES derived from leaf tissue of a tobacco plant transformed with the borage Δ6-desaturase cDNA under transcriptional control of the CaMV 35S promoter (pAN2). Peaks corresponding to methyl linoleate (18:2), methyl γ-linolenate (18:3γ), methyl α-linolenate (18:3α), and methyl octadecatetraenoate (18:4) are indicated.

Fig. 4 is the nucleotide sequence and corresponding amino acid sequence of the oleosin AtS21 cDNA (SEQ ID NO:3).

Fig. 5 is an acidic-base map of the predicted AtS21 protein generated by DNA Strider 1.2.

Fig. 6 is a Kyte-Doolittle plot of the predicted AtS21 protein generated by DNA Strider 1.2.

Fig. 7 is a sequence alignment of oleosins isolated from *Arabidopsis*. Oleosin sequences published or deposited in EMBL, BCM, NCBI databases were aligned to each other using GeneWorks® 2.3. Identical residues are boxed with rectangles. The seven sequences fall into three groups. The first group includes AtS21 (SEQ ID NO:5), X91918 (SEQ ID NO:6) and Z29859 (SEQ ID NO:7). The second group includes X62352 (SEQ ID NO:8) and Ato13 (SEQ ID NO:9). The third group includes X91956 (SEQ ID NO:10) and L40954 (SEQ ID NO:11). Differences in amino acid residues within the same group are indicated by shadows. Ato2/Z54164 is identical to AtS21. Ato13 sequence (Accession No. Z541654 in EMBL database) is actually not disclosed in the EMBL database. The Z54165 Accession number designates the same sequence as Z54164 which is Ato12.

Fig. 8A is a Northern analysis of the AtS21 gene. An RNA gel blot containing ten micrograms of total RNA extracted from *Arabidopsis* flowers (F), leaves (L), roots (R), developing seeds (Se), and developing silique coats (Si) was hybridized with a probe made from the full-length AtS21 cDNA.

Fig. 8B is a Southern analysis of the AtS21 gene. A DNA gel blot containing ten micrograms of genomic DNA digested with BamHI (B), EcoRI (E), HindIII (H), SacI (S), and XbaI (X) was hybridized with a probe made from the full length AtS21 cDNA.

Fig. 9 is the nucleotide sequence of the SacI fragment of AtS21 genomic DNA (SEQ ID NO:12). The promoter and intron sequences are in uppercase. The fragments corresponding to AtS21 cDNA sequence are in lower case. The first ATG codon and a putative TATA box are shadowed. The sequence complementary to 21P primer for PCR amplification is boxed. A putative abscisic acid response element (ABRE) and two 14 bp repeats are underlined.

Fig. 10 is a map of AtS21 promoter/GUS construct (pAN5).

Fig. 11A depicts AtS21/GUS gene expression in *Arabidopsis* bolt and leaves.

Fig. 11B depicts AtS21 GUS gene expression in *Arabidopsis* siliques.

Fig. 11C depicts AtS21 GUS gene expression in *Arabidopsis* developing seeds.

Figs. 11D through 11J depict AtS21 GUS gene expression in *Arabidopsis* developing embryos.

Fig. 11K depicts AtS21/GUS gene expression in *Arabidopsis* root and root hairs of a young seedling.

Fig. 11L depicts AtS21/GUS gene expression in *Arabidopsis* cotyledons and the shoot apex of a five day seedling.

Figs. 11M and 11N depict AtS21/GUS gene expression in *Arabidopsis* cotyledons and the shoot apex of 5-15 day seedlings.

Fig. 12A depicts AtS21/GUS gene expression in tobacco embryos and endosperm.

Fig. 12B depicts AtS21/GUS gene expression in germinating tobacco seeds.

Fig. 12C depicts AtS21/GUS gene expression in a 5 day old tobacco seedling.

Fig. 12D depicts AtS21/GUS gene expression in 5-15 day old tobacco seedlings.

Fig. 13A is a Northern analysis showing AtS21 mRNA levels in developing wild-type *Arabidopsis* seedlings. Lane 1 was loaded with RNA from developing seeds, lane 2 was loaded with RNA from seeds imbibed for 24-48 hours, lane 3: 3 day seedlings; lane 4: 4 day seedlings; lane 5: 5 day seedlings; lane 6: 6 day seedlings; lane 7; 9 day seedlings; lane 8: 12 day seedlings. Probe was labeled AtS21 cDNA. Exposure was for one hour at -80°C.

Fig. 13B is the same blot as Fig. 13A only exposure was for 24 hours at -80°C.

Fig. 13C is the same blot depicted in Figs. 13A and 13B after stripping and hybridization with an *Arabidopsis* tubulin gene probe. The small band in each of lanes 1 and 2 is the remnant of the previous AtS21 probe. Exposure was for 48 hours at -80°C.

Fig. 14 is a graph comparing GUS activities expressed by the AtS21 and 35S promoters. GUS activities expressed by the AtS21 promoter in

developing *Arabidopsis* seeds and leaf are plotted side by side with those expressed by the 35S promoter. The GUS activities expressed by the AtS21 promoter in tobacco dry seed and leaf are plotted on the right side of the figure. GUS activity in tobacco leaf is so low that no column appears. "G-H" denotes globular to heart stage; "H-T" denotes heart to torpedo stage; "T-C" denotes torpedo to cotyledon stage; "Early C" denotes early cotyledon; "Late C" denotes late cotyledon. The standard deviations are listed in Table 2.

Fig. 15A is an RNA gel blot analysis carried out on 5 μg samples of RNA isolated from borage leaf, root, and 12 dpp embryo tissue, using labeled borage Δ6-desaturase cDNA as a hybridization probe.

Fig. 15B depicts a graph corresponding to the Northern analysis results for the experiment shown in Fig. 15A.

Fig. 16A is a graph showing relative legumin RNA accumulation in developing borage embryos based on results of Northern blot.

Fig. 16B is a graph showing relative oleosin RNA accumulation in developing borage embryos based on results of Northern blot.

Fig. 16C is a graph showing relative Δ6-desaturase RNA accumulation in developing borage embryos based on results of Northern blot.

Fig. 17 is a PCR analysis showing the presence of the borage delta 6-desaturase gene in transformed plants of oilseed rape. Lanes 1, 3 and 4

were loaded with PCR reactions performed with DNA from plants transformed with the borage delta 6-desaturase gene linked to the oleosin 5' regulatory region; lane 2: DNA from plant transformed with the borage delta 6-desaturase gene linked to the albumin 5' regulatory region; lanes 5 and 6: DNA from non-transformed plants; lane 7: molecular weight marker (1 kb ladder, Gibco BRL); lane 8: PCR without added template DNA; lane 9: control with DNA from *Agrobacterium tumefaciens* EHA 105 containing the plasmid pAN3 (i.e. the borage delta6-desaturase gene linked to the oleosin 5' regulatory region).

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides isolated nucleic acids encoding 5' regulatory regions from an *Arabidopsis* oleosin gene. In accordance with the present invention, the subject 5' regulatory regions, when operably linked to either a coding sequence of a heterologous gene or a sequence complementary to a native plant gene, direct expression of the coding sequence or complementary sequence in a plant seed. The oleosin 5' regulatory regions of the present invention are useful in the construction of an expression cassette which comprises in the 5' to 3' direction, a subject oleosin 5' regulatory region, a heterologous gene or sequence complementary to a native plant gene under control of the regulatory region and a 3' termination sequence. Such an expression cassette can be incorporated into a variety

of autonomously replicating vectors in order to construct an expression vector.

It has been surprisingly found that plants transformed with the expression vectors of the present invention produce levels of GLA approaching the level found in those few plant species which naturally produce GLA such as evening primrose *(Oenothera biennis)*.

As used herein, the term "cassette" refers to a nucleotide sequence capable of expressing a particular gene if said gene is inserted so as to be operably linked to one or more regulatory regions present in the nucleotide sequence. Thus, for example, the expression cassette may comprise a heterologous coding sequence which is desired to be expressed in a plant seed. The expression cassettes and expression vectors of the present invention are therefore useful for directing seed-specific expression of any number of heterologous genes. The term "seed-specific expression" as used herein, refers to expression in various portions of a plant seed such as the endosperm and embryo.

An isolated nucleic acid encoding a 5' regulatory region from an oleosin gene can be provided as follows. Oleosin recombinant genomic clones are isolated by screening a plant genomic DNA library with a cDNA (or a portion thereof) representing oleosin mRNA. A number of different oleosin cDNAs have been isolated. The methods used to isolate such cDNAs as well as the nucleotide and corresponding amino acid

sequences have been published in Kirik et al. 1986 *Plant Mol. Biol. 31*:413-417; Zou et al. *Plant Mol. Biol. 31*:429-433; Van Rooigen et al. 1992 *Plant Mol. Biol. 18*:1177-1179.

Virtual subtraction screening of a tissue specific library using a random primed polymerase chain (RP-PCR) cDNA probe is another method of obtaining an oleosin cDNA useful for screening a plant genomic DNA library. Virtual subtraction screening refers to a method where a cDNA library is constructed from a target tissue and displayed at a low density so that individual cDNA clones can be easily separated. These cDNA clones are subtractively screened with driver quantities (i.e., concentrations of DNA to kinetically drive the hybridization reaction) of cDNA probes made from tissue or tissues other than the target tissue (i.e. driver tissue). The hybridized plaques represent genes that are expressed in both the target and the driver tissues; the unhybridized plaques represent genes that may be target tissue-specific or low abundant genes that can not be detected by the driver cDNA probe. The unhybridized cDNAs are selected as putative target tissue-specific genes and further analyzed by one-pass sequencing and Northern hybridization.

Random primed PCR (RP-PCR) involves synthesis of large quantities of cDNA probes from a trace amount of cDNA template. The method combines the amplification power of PCR with the representation

of random priming to simultaneously amplify and label double-stranded cDNA in a single tube reaction.

Methods considered useful in obtaining oleosin genomic recombinant DNA are provided in Sambrook et al. 1989, in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, NY, for example, or any of the myriad of laboratory manuals on recombinant DNA technology that are widely available. To determine nucleotide sequences, a multitude of techniques are available and known to the ordinarily skilled artisan. For example, restriction fragments containing an oleosin regulatory region can be subcloned into the polylinker site of a sequencing vector such as pBluescript (Stratagene). These pBluescript subclones can then be sequenced by the double-stranded dideoxy method (Chen and Seeburg, 1985, *DNA* 4:165).

In a preferred embodiment, the oleosin regulatory region comprises nucleotides 1-1267 of Fig. 9 (SEQ ID NO:12). Modifications to the oleosin regulatory region as set forth in SEQ ID NO:12 which maintain the characteristic property of directing seed-specific expression, are within the scope of the present invention. Such modifications include insertions, deletions and substitutions of one or more nucleotides.

The 5' regulatory region of the present invention can be derived from restriction endonuclease or exonuclease digestion of an oleosin genomic clone. Thus, for example, the known nucleotide or amino acid

sequence of the coding region of an isolated oleosin gene (e.g. Fig. 7) is aligned to the nucleic acid or deduced amino acid sequence of an isolated oleosin genomic clone and 5' flanking sequence (i.e., sequence upstream from the translational start codon of the coding region) of the isolated oleosin genomic clone located.

The oleosin 5' regulatory region as set forth in SEQ ID NO:12 (nucleotides 1-1267 of Fig. 9) may be generated from a genomic clone having either or both excess 5' flanking sequence or coding sequence by exonuclease III-mediated deletion. This is accomplished by digesting appropriately prepared DNA with exonuclease III (exoIII) and removing aliquots at increasing intervals of time during the digestion. The resulting successively smaller fragments of DNA may be sequenced to determine the exact endpoint of the deletions. There are several commercially available systems which use exonuclease III (exoIII) to create such a deletion series, e.g. Promega Biotech, "Erase-A-Base" system. Alternatively, PCR primers can be defined to allow direct amplification of the subject 5' regulatory regions.

Using the same methodologies, the ordinarily skilled artisan can generate one or more deletion fragments of nucleotides 1-1267 as set forth in SEQ ID NO:12. Any and all deletion fragments which comprise a contiguous portion of nucleotides set forth in SEQ ID NO:12 and which retain the capacity to

direct seed-specific expression are contemplated by the present invention.

The identification of oleosin 5' regulatory sequences which direct seed-specific expression comprising nucleotides 1-1267 of SEQ ID NO:12 and modifications or deletion fragments thereof, can be accomplished by transcriptional fusions of specific sequences with the coding sequences of a heterologous gene, transfer of the chimeric gene into an appropriate host, and detection of the expression of the heterologous gene. The assay used to detect expression depends upon the nature of the heterologous sequence. For example, reporter genes, exemplified by chloramphenicol acetyl transferase and β-glucuronidase (GUS), are commonly used to assess transcriptional and translational competence of chimeric constructions. Standard assays are available to sensitively detect the reporter enzyme in a transgenic organism. The β-glucuronidase (GUS) gene is useful as a reporter of promoter activity in transgenic plants because of the high stability of the enzyme in plant cells, the lack of intrinsic β-glucuronidase activity in higher plants and availability of a quantitative fluorimetric assay and a histochemical localization technique. Jefferson et al. (1987 *EMBO J 6*:3901) have established standard procedures for biochemical and histochemical detection of GUS activity in plant tissues. Biochemical assays are performed by mixing plant tissue lysates with 4-methylumbelliferyl-β-D-glucuronide, a fluorimetric substrate for GUS, incubating one hour at 37°C, and

then measuring the fluorescence of the resulting 4-methyl-umbelliferone. Histochemical localization for GUS activity is determined by incubating plant tissue samples in 5-bromo-4-chloro-3-indolyl-glucuronide (X-Gluc) for about 18 hours at 37°C and observing the staining pattern of X-Gluc. The construction of such chimeric genes allows definition of specific regulatory sequences and demonstrates that these sequences can direct expression of heterologous genes in a seed-specific manner.

Another aspect of the invention is directed to expression cassettes and expression vectors (also termed herein "chimeric genes") comprising a 5' regulatory region from an oleosin gene which directs seed specific expression operably linked to the coding sequence of a heterologous gene such that the regulatory element is capable of controlling expression of the product encoded by the heterologous gene. The heterologous gene can be any gene other than oleosin. If necessary, additional regulatory elements or parts of these elements sufficient to cause expression resulting in production of an effective amount of the polypeptide encoded by the heterologous gene are included in the chimeric constructs.

Accordingly, the present invention provides chimeric genes comprising sequences of the oleosin 5' regulatory region that confer seed-specific expression which are operably linked to a sequence encoding a heterologous gene such as a lipid metabolism enzyme.

Examples of lipid metabolism genes useful for practicing the present invention include lipid desaturases such as Δ6-desaturases, Δ12-desaturases, Δ15-desaturases and other related desaturases such as stearoyl-ACP desaturases, acyl carrier proteins (ACPs), thioesterases, acetyl transacylases, acetyl-coA carboxylases, ketoacyl-synthases, malonyl transacylases, and elongases. Such lipid metabolism genes have been isolated and characterized from a number of different bacteria and plant species. Their nucleotide coding sequences as well as methods of isolating such coding sequences are disclosed in the published literature and are widely available to those of skill in the art.

In particular, the Δ6-desaturase genes disclosed in U.S. Patent No. 5,552,306 and applicants' copending U.S. Application Serial No. 08/366,779 filed December 30, 1994 and incorporated herein by reference, are contemplated as lipid metabolism genes particularly useful in the practice of the present invention.

The chimeric genes of the present invention are constructed by ligating a 5' regulatory region of a oleosin genomic DNA to the coding sequence of a heterologous gene. The juxtaposition of these sequences can be accomplished in a variety of ways. In a preferred embodiment the order of the sequences, from 5' to 3', is an oleosin 5' regulatory region (including a promoter), a coding sequence, and a

termination sequence which includes a polyadenylation site.

Standard techniques for construction of such chimeric genes are well known to those of ordinary skill in the art and can be found in references such as Sambrook et al.(1989). A variety of strategies are available for ligating fragments of DNA, the choice of which depends on the nature of the termini of the DNA fragments. One of ordinary skill in the art recognizes that in order for the heterologous gene to be expressed, the construction requires promoter elements and signals for efficient polyadenylation of the transcript. Accordingly, the oleosin 5' regulatory region that contains the consensus promoter sequence known as the TATA box can be ligated directly to a promoterless heterologous coding sequence.

The restriction or deletion fragments that contain the oleosin TATA box are ligated in a forward orientation to a promoterless heterologous gene such as the coding sequence of β-glucuronidase (GUS). The skilled artisan will recognize that the subject oleosin 5' regulatory regions can be provided by other means, for example chemical or enzymatic synthesis. The 3' end of a heterologous coding sequence is optionally ligated to a termination sequence comprising a polyadenylation site, exemplified by, but not limited to, the nopaline synthase polyadenylation site, or the octopine T-DNA gene 7 polyadenylation site. Alternatively, the polyadenylation site can be provided by the heterologous gene.

The present invention also provides methods of increasing levels of heterologous genes in plant seeds. In accordance with such methods, the subject expression cassettes and expression vectors are introduced into a plant in order to effect expression of a heterologous gene. For example, a method of producing a plant with increased levels of a product of a fatty acid synthesis or lipid metabolism gene is provided by transforming a plant cell with an expression vector comprising an oleosin 5' regulatory region operably linked to a fatty acid synthesis or lipid metabolism gene and regenerating a plant with increased levels of the product of said fatty acid synthesis or lipid metabolism gene.

Another aspect of the present invention provides methods of reducing levels of a product of a gene which is native to a plant which comprises transforming a plant cell with an expression vector comprising a subject oleosin regulatory region operably linked to a nucleic acid sequence which is complementary to the native plant gene. In this manner, levels of endogenous product of the native plant gene are reduced through the mechanism known as antisense regulation. Thus, for example, levels of a product of a fatty acid synthesis gene or lipid metabolism gene are reduced by transforming a plant with an expression vector comprising a subject oleosin 5' regulatory region operably linked to a nucleic acid sequence which is complementary to a nucleic acid

sequence coding for a native fatty acid synthesis or lipid metabolism gene.

The present invention also provides a method of cosuppressing a gene which is native to a plant which comprises transforming a plant cell with an expression vector comprising a subject oleosin 5' regulatory region operably linked to a nucleic acid sequence coding for the native plant gene. In this manner, levels of endogenous product of the native plant gene are reduced through the mechanism known as cosuppression. Thus, for example, levels of a product of a fatty acid synthesis gene or lipid metabolism gene are reduced by transforming a plant with an expression vector comprising a subject oleosin 5' regulatory region operably linked to a nucleic acid sequence coding for a native fatty acid synthesis or lipid metabolism gene native to the plant. Although the exact mechanism of cosuppression is not completely understood, one skilled in the art is familiar with published works reporting the experimental conditions and results associated with cosuppression (Napoli et al. 1990 *The Plant Cell 2*:270-289; Van der Krol 1990 *The Plant Cell 2*:291-299.

To provide regulated expression of the heterologous or native genes, plants are transformed with the chimeric gene constructions of the invention. Methods of gene transfer are well known in the art. The chimeric genes can be introduced into plants by leaf disk transformation-regeneration procedure as described by Horsch et al. 1985 *Science 227*:1229.

Other methods of transformation such as protoplast culture (Horsch et al. 1984 *Science 223*:496, DeBlock et al. 1984 *EMBO J. 2*:2143, Barton et al. 1983, *Cell 32*:1033) can also be used and are within the scope of this invention. In a preferred embodiment, plants are transformed with *Agrobacterium*-derived vectors such as those described in Klett et al. (1987) *Annu. Rev. Plant Physiol. 38*:467. Other well-known methods are available to insert the chimeric genes of the present invention into plant cells. Such alternative methods include biolistic approaches (Klein et al. 1987 *Nature 327*:70), electroporation, chemically-induced DNA uptake, and use of viruses or pollen as vectors.

When necessary for the transformation method, the chimeric genes of the present invention can be inserted into a plant transformation vector, e.g. the binary vector described by Bevan, M. 1984 *Nucleic Acids Res. 12*:8711-8721. Plant transformation vectors can be derived by modifying the natural gene transfer system of *Agrobacterium tumefaciens*. The natural system comprises large Ti (tumor-inducing)-plasmids containing a large segment, known as T-DNA, which is transferred to transformed plants. Another segment of the Ti plasmid, the *vir* region, is responsible for T-DNA transfer. The T-DNA region is bordered by terminal repeats. In the modified binary vectors, the tumor inducing genes have been deleted and the functions of the *vir* region are utilized to transfer foreign DNA bordered by the T-DNA border sequences. The T-region also contains a selectable

marker for antibiotic resistance, and a multiple cloning site for inserting sequences for transfer. Such engineered strains are known as "disarmed" *A. tumefaciens* strains, and allow the efficient transfer of sequences bordered by the T-region into the nuclear genome of plants.

Surface-sterilized leaf disks and other susceptible tissues are inoculated with the "disarmed" foreign DNA-containing *A. tumefaciens*, cultured for a number of days, and then transferred to antibiotic-containing medium. Transformed shoots are then selected after rooting in medium containing the appropriate antibiotic, and transferred to soil. Transgenic plants are pollinated and seeds from these plants are collected and grown on antibiotic medium.

Expression of a heterologous or reporter gene in developing seeds, young seedlings and mature plants can be monitored by immunological, histochemical or activity assays. As discussed herein, the choice of an assay for expression of the chimeric gene depends upon the nature of the heterologous coding region. For example, Northern analysis can be used to assess transcription if appropriate nucleotide probes are available. If antibodies to the polypeptide encoded by the heterologous gene are available, Western analysis and immunohistochemical localization can be used to assess the production and localization of the polypeptide. Depending upon the heterologous gene, appropriate biochemical assays can be used. For example,

acetyltransferases are detected by measuring acetylation of a standard substrate. The expression of a lipid desaturase gene can be assayed by analysis of fatty acid methyl esters (FAMES).

Another aspect of the present invention provides transgenic plants or progeny of these plants containing the chimeric genes of the invention. Both monocotyledonous and dicotyledonous plants are contemplated. Plant cells are transformed with the chimeric genes by any of the plant transformation methods described above. The transformed plant cell, usually in the form of a callus culture, leaf disk, explant or whole plant (via the vacuum infiltration method of Bechtold et al. 1993 *C.R. Acad. Sci. Paris*, *316*:1194-1199) is regenerated into a complete transgenic plant by methods well-known to one of ordinary skill in the art (e.g. Horsch et al. 1985 *Science 227*:1129). In a preferred embodiment, the transgenic plant is sunflower, cotton, oil seed rape, maize, tobacco, *Arabidopsis*, peanut or soybean. Since progeny of transformed plants inherit the chimeric genes, seeds or cuttings from transformed plants are used to maintain the transgenic line.

The following examples further illustrate the invention.

EXAMPLE 1

Isolation of Membrane-Bound Polysomal RNA and Construction of Borage cDNA Library

Membrane-bound polysomes were isolated from borage seeds 12 days post pollination (12 DPP)using the protocol established for peas by Larkins and Davies (1975 *Plant Phys. 55:* 749-756). RNA was extracted from the polysomes as described by Mechler (1987 *Methods in Enzymology 152:* 241-248, Academic Press). Poly-A^+ RNA was isolated from the membrane bound polysomal RNA using Oligotex-dT™ beads (Qiagen).

Corresponding cDNA was made using Stratagene's ZAP cDNA synthesis kit. The cDNA library was constructed in the lambda ZAP II vector (Stratagene) using the lambda ZAP II kit. The primary library was packaged with Gigapack II Gold packaging extract (Stratagene).

EXAMPLE 2

Isolation of a Δ-6 Desaturase cDNA from Borage

Hybridization protocol

The amplified borage cDNA library was plated at low density (500 pfu on 150 mm petri dishes). Highly prevalent seed storage protein cDNAs were reduced (subtracted from the total cDNAs) by screening with the corresponding cDNAs.

Hybridization probes for screening the borage cDNA library were generated by using random primed DNA synthesis as described by Ausubel et al (1994 Current Protocols in Molecular Biology, Wiley Interscience, N.Y.) and corresponded to previously identified abundantly expressed seed storage protein cDNAs. Unincorporated nucleotides were removed by use of a G-50 spin column (Boehringer Manheim). Probe was denatured for hybridization by boiling in a water bath for 5 minutes, then quickly cooled on ice. Nitrocellulose filters carrying fixed recombinant bacteriophage were prehybridized at 60°C for 2-4 hours in hybridization solution [4X SET (600 mM NaCl, 80 mM Tris-HCl, 4 mM Na_2EDTA; pH 7.8), 5X Denhardt's reagent (0.1% bovine serum albumin, 0.1% Ficoll, and 0.1% polyvinylpyrolidone), 100 μg/ml denatured salmon sperm DNA, 50 μg/ml polyadenine and 10 ug/ml polycytidine]. This was replaced with fresh hybridization solution to which denatured radioactive probe (2 ng/ml hybridization solution) was added. The filters were incubated at 60°C with agitation overnight. Filters

were washed sequentially in 4X, 2X, and 1X SET (150 mM NaCl, 20 mM Tris-HCl, 1 mM Na_2EDTA; pH7.8) for 15 minutes each at 60°C. Filters were air dried and then exposed to X-ray film for 24 hours with intensifying screens at -80°C.

Non-hybridizing plaques were excised using Stratagene's excision protocol and reagents. Resulting bacterial colonies were used to inoculate liquid cultures and were either sequenced manually or by an ABI automated sequencer.

Random Sequencing of cDNAs from a Borage Seed 12 (DPP) Membrane-Bound Polysomal Library

Each cDNA corresponding to a non-hybridizing plaque was sequenced once and a sequence tag generated from 200-300 base pairs. All sequencing was performed by cycle sequencing (Epicentre). Over 300 expressed sequence tags (ESTs) were generated. Each sequence tag was compared to the GenBank database using the BLAST algorithm (Altschul et al. 1990 *J. Mol. Biol. 215*:403-410). A number of lipid metabolism genes, including the Δ6-desaturase were identified.

Database searches with the cDNA clone designated mbp-65 using BLASTX with the GenBank database resulted in a significant match to the previously isolated *Synechocystis* Δ6-desaturase. It was determined however, that mbp-65 was not a full length cDNA. A full length cDNA was isolated using mbp-65 to screen the borage membrane-bound polysomal library. The resultant clone was designated pAN1 and the cDNA insert of pAN1 was sequenced by the cycle

sequencing method. The amino acid sequence deduced from the open reading frame (Fig. 1, SEQ ID NO:1) was compared to other known desaturases using Geneworks (IntelligGenetics) protein alignment program. This alignment indicated that the cDNA insert of pAN1 was the borage Δ6-desaturase gene.

The resulting dendrogram (Figure 2) shows that Δ^{15}-desaturases and Δ^{12}-desaturases comprise two groups. The newly isolated borage sequence and the previously isolated *Synechocystis* Δ^{6}-desaturase (U.S. Patent No. 5,552,306) formed a third distinct group. A comparison of amino acid motifs common to desaturases and thought to be involved catalytically in metal binding illustrates the overall similarity of the protein encoded by the borage gene to desaturases in general and the *Synechocystis* Δ^{6}-desaturase in particular (Table 1). At the same time, comparison of the motifs in Table 1 indicates definite differences between this protein and other plant desaturases. Furthermore, the borage sequence is also distinguished from known plant membrane associated fatty acid desaturases by the presence of a heme binding motif conserved in cytochrome b_5 proteins (Schmidt et al. 1994 *Plant Mol. Biol. 26*:631-642)(Figure 1). Thus, while these results clearly suggested that the isolated cDNA was a borage Δ^{6}-desaturase gene, further confirmation was necessary. To confirm the identity of the borage Δ6-desaturase cDNA, the cDNA insert from pAN1 was cloned into an expression cassette for stable expression. The vector pBI121 (Jefferson et al. 1987

EMBO J. 6:3901-3907) was prepared for ligation by digestion with BamHI and EcoICR I (an isoschizomer of SacI which leaves blunt ends; available from Promega) which excises the GUS coding region leaving the 35S promoter and NOS terminator intact. The borage Δ^6-desaturase cDNA was excised from the recombinant plasmid (pAN1) by digestion with BamHI and XhoI. The XhoI end was made blunt by performing a fill-in reaction catalyzed by the Klenow fragment of DNA polymerase I. This fragment was then cloned into the BamHI/EcoICR I sites of pBI121.1, resulting in the plasmid pAN2.

TABLE 1

COMPARISON OF COMMON AMINO ACID MOTIFS IN MEMBRANE-BOUND DESATURASES

Desaturase	Lipid Box	Metal Box 1	Metal Box 2
Borage Δ^{6}	WIGHDAGH (SEQ. ID. NO:15)	HNAHH (SEQ. ID. NO:21)	FQIEHH (SEQ. ID. NO:29)
Synechocystis Δ^{6}	NVGHDANH (SEQ. ID. NO:16)	HNYLHH (SEQ. ID. NO:22)	HQVTHH (SEQ. ID. NO:30)
Arab. chloroplast Δ^{15}	VLGHDCGH (SEQ. ID. NO:17)	HRTHH (SEQ. ID. NO:23)	HVIHH (SEQ. ID. NO:31)
Rice Δ^{15}	VLGHDCGH (SEQ. ID. NO:17)	HRTHH (SEQ. ID. NO:23)	HVIHH (SEQ. ID. NO:31)
Glycine chloroplast Δ^{15}	VLGHDCGH (SEQ. ID. NO:17)	HRTHH (SEQ. ID. NO:23)	HVIHH (SEQ. ID. NO:31)
Arab. fad3 (Δ^{15})	VLGHDCGH (SEQ. ID. NO:17)	HRTHH (SEQ. ID. NO:23)	HVIHH (SEQ. ID. NO:31)
Brassica fad 3 (Δ^{15}	VLGHDCGH (SEQ. ID. NO:17)	HRTHH (SEQ. ID. NO:23)	HVIHH (SEQ. ID. NO:31)
Borage Δ^{12} (P1-81)*	VIAHECGH (SEQ. ID. NO:18)	HRRHH (SEQ. ID. NO:24)	HVAHH (SEQ. ID. NO:32)
Arab. fad2 (Δ^{12})	VIAHECGH (SEQ. ID. NO:18)	HRRHH (SEQ. ID. NO:24)	HVAHH (SEQ. ID. NO:32)
Arab. chloroplast Δ^{12}	VIGHDCAH (SEQ. ID. NO:19)	HDRHH (SEQ. ID. NO:25)	HIPHH (SEQ. ID. NO:33)
Glycine plastid Δ^{12}	VIGHDCAH (SEQ. ID. NO:19)	HDRHH (SEQ. ID. NO:25)	HIPHH (SEQ. ID. NO:33)
Spinach plastidial n-6	VIGHDCAH (SEQ. ID. NO:19)	HDQHH (SEQ. ID. NO:26)	HIPHH (SEQ. ID. NO:33)
Synechocystis Δ^{12}	VVGHDCGH (SEQ. ID. NO:20)	HDHHH (SEQ. ID. NO:27)	HIPHH (SEQ. ID. NO:33)
Anabaena Δ^{12}	VLGHDCGH (SEQ. ID. NO:17)	HNHHH (SEQ. ID. NO:28)	HVPHH (SEQ. ID. NO:34)

*P1-81 is a full length cDNA which was identified by EST analysis and shows high similarity to the Arbidopsis Δ12 desaturase (fad2)

EXAMPLE 3

Production of Transgenic Plants and Preparation and Analysis of Fatty Acid Methyl Esters (FAMEs)

The expression plasmid, pAN2 was used to transform tobacco (*Nicotiana tabacum* cv. *xanthi*) via *Agrobacterium tumefaciens* according to standard procedures (Horsch, et al. 1985 *Science* 227:1229-1231; Bogue et al. 1990 *Mol. Gen. Genet. 221:*49-57) except that the initial transformants were selected on 100 μg/ml kanamycin.

Tissue from transgenic plants was frozen in liquid nitrogen and lyophilized overnight. FAMEs were prepared as described by Dahmer, et al. (1989) *J. Amer. Oil. Chem. Soc. 66:* 543-548. In some cases, the solvent was evaporated again, and the FAMEs were resuspended in ethyl acetate and extracted once with deionized water to remove any water soluble contaminants. FAMEs were analyzed using a Tracor-560 gas liquid chromatograph as previously described (Reddy et al. 1996 *Nature Biotech. 14*:639-642).

As shown in Figure. 3, transgenic tobacco leaves containing the borage cDNA produced both GLA and octadecatetraenoic acid (OTA) (18:4 Δ6,9,12,15). These results thus demonstrate that the isolated cDNA encodes a borage Δ6-desaturase.

EXAMPLE 4

Expression of Δ6-desaturase in Borage

The native expression of Δ6-desaturase was examined by Northern Analysis of RNA derived from borage tissues. RNA was isolated from developing borage embryos following the method of Chang et al. 1993 *Plant Mol. Biol. Rep. 11*:113-116. RNA was electrophoretically separated on formaldehyde-agarose gels, blotted to nylon membranes by capillary transfer, and immobilized by baking at 80°C for 30 minutes following standard protocols (Brown T., 1996 in *Current Protocols in Molecular Biology,* eds. Auselbel, et al. [Greene Publishing and Wiley-Interscience, New York] pp. 4.9.1-4.9.14.). The filters were preincubated at 42°C in a solution containing 50% deionized formamide, 5X Denhardt's reagent, 5X SSPE (900 mM NaCl; 50mM Sodium phosphate, pH7.7; and 5 mM EDTA), 0.1% SDS, and 200 μg/ml denatured salmon sperm DNA. After two hours, the filters were added to a fresh solution of the same composition with the addition of denatured radioactive hybridization probe. In this instance, the probes used were borage legumin cDNA (Fig. 16A), borage oleosin cDNA (Fig. 16B), and borage Δ6-desaturase cDNA (pAN1, Example 2)(Fig. 16C). The borage legumin and oleosin cDNAs were isolated by EST cloning and identified by comparison to the GenBank database using the BLAST algorithm as described in Example 2. Loading variation was corrected by normalizing to

levels of borage EF1α mRNA. EF1α mRNA was identified by correlating to the corresponding cDNA obtained by the EST analysis described in Example 2. The filters were hybridized at 42°C for 12-20 hours, then washed as described above (except that the temperature was 65°C), air dried, and exposed to X-ray film.

As depicted in Figs. 15A and 15B, Δ6-desaturase is expressed primarily in borage seed. Borage seeds reach maturation between 18-20 days post pollination (dpp). Δ6-desaturase mRNA expression occurs throughout the time points collected (8-20 dpp), but appears maximal from 10-16 days post pollination. This expression profile is similar to that seen for borage oleosin and 12S seed storage protein mRNAs (Figs. 16A, 16B, and 16C).

EXAMPLE 5

Isolation and Characterization of a Novel Oleosin cDNA

The oleosin cDNA (AtS21) was isolated by virtual subtraction screening of an *Arabidopsis* developing seed cDNA library using a random primed polymerase chain reaction (RP-PCR) cDNA probe derived from root tissue.

RNA PREPARATION

Arabidopsis thaliana Landsberg *erecta* plants were grown under continuous illumination in a vermiculite/soil mixture at ambient temperature (22°C). Siliques 2-5 days after flowering were dissected to separately collect developing seeds and silique coats. Inflorescences containing initial flower buds and fully opened flowers, leaves, and whole siliques one or three days after flowering were also collected. Roots were obtained from seedlings that had been grown in Gamborg B_5 liquid medium (GIBCO BRL) for two weeks. The seeds for root culture were previously sterilized with 50% bleach for five minutes and rinsed with water extensively. All tissues were frozen in liquid nitrogen and stored at -80°C until use. Total RNAs were isolated following a hot phenol/SDS extraction and LiCl precipitation protocol (Harris et al. 1978 *Biochem.* *17*:3251-3256; Galau et al. 1981 *J. Biol. Chem.* *256*:2551-2560). Poly A+ RNA was isolated using oligo dT column chromatography according to manufacturers' protocols (PHARMACIA or

STRATAGENE) or using oligotex-dT latex particles (QIAGEN).

Construction of tissue-specific cDNA libraries

Flower, one day silique, three day silique, leaf, root, and developing seed cDNA libraries were each constructed from 5 µg poly A+ RN using the ZAP cDNA synthesis kit (Stratagene). cDNAs were directionally cloned into the EcoRI and XhoI sites of pBluescript SK(-) in the λ-ZAPII vector (Short et al. 1988 *Nucleic Acids Res.* 16:7583-7600). Nonrecombinant phage plaques were identified by blue color development on NZY plates containing X-gal (5 bromo-4-chloro-3-indoyl-β-D-galactopyranoside) and IPTG (isopropyl-1-thio-β-D-galactopyranoside). The nonrecombinant backgrounds for the flower, one day silique, three day silique, leaf, root, and developing seed cDNA libraries were 2.8%, 2%m 3.3%, 6.5%, 2.5%, and 1.9% respectively.

Random priming DNA labeling

The cDNA inserts of isolated clones (unhybridized cDNAs) were excised by EcoRI/XhoI double digestion and gel-purified for random priming labeling. Klenow reaction mixture contained 50 ng DNA templates, 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 7.5 mM DTT, 50 uM each of dCTP, dGTP, and dTTP, 10 uM hexamer random primbers (Boehringer Mannheim), 50 µCi α-32 P-dATP, 3000 Ci/mmole, 10 mCi/ml (DuPont), and 5 units of DNA polymerase I Klenow fragment (New England

Biolabs). The reactions were carried out at 37°C for one hour. Aliquots of diluted reaction mixtures were used for TCA precipitation and alkaline denaturing gel analysis. Hybridization probes were labeled only with Klenow DNA polymerase and the unincorporated dNTPs were removed using Sephadex R G-50 spin columns (Boehringer Mannheim).

Random Primed PCR

Double-stranded cDNA was synthesized from poly A+ RNA isolated from *Arabidopsis* root tissue using the cDNA Synthesis System (GIBCO BRL) with oligo dT12-18 as primers. cDNAs longer than 300 bp were enriched by Sephacryl S-400 column chromatography (Stratagene). Fractionated cDNAs were used as templates for RP-PCR labeling. The reaction contained 10 mM Tris-HCl, ph 9.0, 50 mM KCl, 0.1% Triton X-100, 2 mM MgCl2, 5 units *Taq* DNA polymeras (PROMEGA), 200 µM dCTP, cGTP, and dTTP, and different concentrations of hexamer random primers α-32P dATP, 800 mCi/mmole, 10 mCi/ml (DuPont), and cold dATP in a final volume of 25 µl. After an initial 5 minutes at 95°C, different reactions were run through different programs to optimize RP-PCR cDNA conditions. Unless otherwise indicated, the following program was used for most RP-PCR cDNA probe labeling: 95°C/5 minutes, then 40 cycles of 95°C 30 seconds, 18°C/1 second, ramp to 30°C at a rate of 0.1°C/second. 72°C/1 minute. RP-PCR products were phenol/chloroform extracted and ethanol

precipitated or purified by passing through Sephadex G-50 spin columns (Boehringer Mannheim).

Clone blot virtual subtraction

Mass excision of λ-ZAP cDNA libraries was carried out by co-infecting XL1-Blue MRF' host cells with recombinant phage from the libraries and ExAssist helper phage (STRATAGENE). Excised phagemids were rescued by SOLR cells. Plasmid DNAs were prepared by boiling mini-prep method (Holmes et al. 1981 *Anal. Biochem. 114*:193-197) from randomly isolated clones. cDNA inserts were excised by EcoRI and XhoI double digestion, and resolved on 1% agarose gels. The DNAs were denatured in 0.5 N NaOH and 1.5 m NaCl for 45 minutes, neutralized in 0.5 M Tris-HCl, pH 8.0, and 1.5 M NaCl for 45 minutes, and then transferred by blotting to nylon membranes (Micron Separations, Inc.) in 10X SSC overnight. After one hour prehybridization at 65°C, root RP-cDNA probe was added to the same hybridization buffer containing 1% bovine albumin fraction V (Sigma), 1 mM EDTA, 0.5 M NaHPO4, pH 7.2, 7% SDS. The hybridization continued for 24 hours at 65°C. The filters were washed in 0.5% bovine albumin, 1 mM EDTA, 40 mM NaHPO4, pH 7.2, 5% SDS for ten minutes at room temperature, and 3 x 10 minutes in 1 mM EDTA, 40 mM NaHPO4, pH 7.2, 1% SDS at 65°C. Autoradiographs were exposed to X-ray films (Kodak) for two to five days at -80°C.

Hybridization of resulting blots with root RP-PCR probes "virtually subtracted" seed cDNAs shared

with the root mRNA population. The remaining seed cDNAs representing putative seed-specific cDNAs, including those encoding oleosins, were sequenced by the cycle sequencing method, thereby identifying AtS21 as an oleosin cDNA clone.

Sequence analysis of AtS21

The oleosin cDNA is 834 bp long including an 18 bp long poly A tail (Fig. 4, SEQ ID NO:2) It has high homology to other oleosin genes from *Arabidopsis* as well as from other species. Recently, an identical oleosin gene has been reported (Zou, et al., 1996, *Plant Mol.Biol. 31*:429-433). The predicted protein is 191 amino acids long with a highly hydrophobic middle domain flanked by a hydrophilic domain on each side. The existence of two upstream in frame stop codons and the similarity to other oleosin genes indicate that this cDNA is full-length. Since there are two in frame stop codons just upstream of the first ATG, this cDNA is considered to be a full length cDNA (Figure 4, SEQ ID NO:2). The predicted protein has three distinctive domains based on the distribution of its amino acid residues. Both the N-terminal and C-terminal domains are rich in charged residues while the central domain is absolutely hydrophobic (Figure 5). As many as 20 leucine residues are located in the central domain and arranged as repeats with one leucine occurring every 7-10 residues. Other non-polar amino acid residues are also clustered in the central domain making this domain absolutely hydrophobic (Figure 6).

Extensive searches of different databases using both AtS21 cDNA and its predicted protein sequence identified oleosins from carrot, maize, cotton, rapeseed, *Arabidopsis*, and other plant species. The homology is mainly restricted to the central hydrophobic domain. Seven *Arabidopsis* oleosin sequences were found. AtS21 represents the same gene as Z54164 which has a few more bases in the 5' untranslated region. The seven *Arabidopsis* oleosin sequences available so far were aligned to each other (Figure 7). The result suggested that the seven sequences fall into three groups. The first group includes AtS21 (SEQ ID NO:5), X91918 (SEQ ID NO:6), and the partial sequence Z29859 (SEQ ID NO:7). Since X91918 (SEQ ID NO:6) has only its last residue different from AtS21 (SEQ ID NO:5), and since Z29859 (SEQ ID NO:7) has only three amino acid residues which are different from AtS21 (SEQ ID NO:5), all three sequences likely represent the same gene. The two sequences of the second group, X62352 (SEQ ID NO:8) and Ato13 (SEQ ID NO:9), are different in both sequence and length. Thus, there is no doubt that they represent two independent genes. Like the first group, the two sequences of the third group, X91956 (SEQ ID NO:10) and L40954 (SEQ ID NO:11), also have only three divergent residues which may be due to sequence errors. Thus, X91956 (SEQ ID NO:10) and L40954 (SEQ ID NO:11) likely represent the same gene. Unlike all the other oleosin sequences which were predicted from cDNA sequences, X62352 (SEQ ID NO:8)

was deduced from a genomic sequence (Van Rooigen et al. 1992 *Plant Mol. Biol.* *18*:1177-1179). In conclusion, four different *Arabidopsis* oleosin genes have been identified so far, and they are conserved only in the middle of the hydrophobic domain.

Northern Analysis

In order to characterize the expression pattern of the native AtS21 gene, Northern analysis was performed as described in Example 4 except that the probe was the AtS21 cDNA (pAN1 insert) labeled with ^{32}P-dATP to a specific activity of 5 x 10^8 cpm/ug.

Results indicated that the AtS21 gene is strongly expressed in developing seeds and weakly expressed in silique coats (Figure 8A). A much larger transcript, which might represent unprocessed AtS21 pre-mRNA, was also detected in developing seed RNA. AtS21 was not detected in flower, leaf, root (Figure 8A), or one day silique RNAs. A different Northern analysis revealed that AtS21 is also strongly expressed in imbibed germinating seeds (Figs. 13A and 13B)

EXAMPLE 6

Characterization of Oleosin Genomic Clones and Isolation of Oleosin Promoter

Genomic clones were isolated by screening an *Arabidopsis* genomic DNA library using the full length cDNA (AtS21)as a probe. Two genomic clones were mapped by restriction enzyme digestion followed by Southern hybridization using the 5' half of the cDNA cleaved by SacI as a probe. A 2 kb SacI fragment was subcloned and sequenced (Fig. 9, SEQ ID NO:35). Two regions of the genomic clone are identical to the cDNA sequence. A 395 bp intron separates the two regions.

The copy number of AtS21 gene in the *Arabidopsis* genome was determined by genomic DNA Southern hybridization following digestion with the enzymes BamHI, EcoRI, HindIII, SacI and XbaI, using the full length cDNA as a probe (Figure 8B). A single band was detected in all the lanes except SacI digestion where two bands were detected. Since the cDNA probe has an internal SacI site, these results indicated that AtS21 is a single copy gene in the *Arabidopsis* genome. Since it has been known that Arabidopsis genome contains different isoforms of oleosin genes, this Southern analysis also demonstrates that the different oleosin isoforms of *Arabidopsis* are divergent at the DNA sequence level.

Two regions, separated by a 395 bp intron, of the genomic DNA fragment are identical to AtS21 cDNA sequence. Database searches using the 5' promoter sequence upstream of AtS21 cDNA sequence did

not identify any sequence with significant homology. Furthermore, the comparison of AtS21 promoter sequence with another *Arabidopsis* oleosin promoter isolated previously (Van Rooijen, et al., 1992) revealed little similarity. The AtS21 promoter sequence is rich in A/T bases, and contains as many as 44 direct repeats ranging from 10 bp to 14 bp with only one mismatch allowed. Two 14 bp direct repeats, and a putative ABA response element are underlined in Figure 9.

EXAMPLE 7

Construction of AtS21 Promoter/GUS Gene Expression Cassette and Expression Patterns in Transgenic *Arabidopsis* and Tobacco

Construction of AtS21 promoter/GUS gene expression cassette

The 1267 bp promoter fragment starting from the first G upstream of the ATG codon of the genomic DNA fragment was amplified using PCR and fused to the GUS reporter gene for analysis of its activity. The promoter fragment of the AtS21 genomic clone was amplified by PCR using the T7 primer GTAATACGACTCACTATAGGGC (SEQ ID NO:13) and the 21P primer GGGGATCCTATACTAAAACTATAGAGTAAAGG (SEQ ID NO:14) complementary to the 5' untranslated region upstream of the first ATG codon (Figure 9). A BamHI cloning site was introduced by the 21P primer. The amplified fragment was cloned into the BamHI and SacI sites of pBluescript KS (Stratagene). Individual clones were sequenced to check possible PCR mutations as well as the orientation of their inserts. The correct clone was digested with BamHI and HindIII, and the excised promoter fragment (1.3 kb) was cloned into the corresponding sites of pBI101.1 (Jefferson, R.A. 1987a, *Plant Mol. Biol. Rep.* 5:387-405; Jefferson et al., 1987b, *EMBO J.* 6:3901-3907) upstream of the GUS gene. The resultant plasmid was designated pAN5 (Fig. 10). The AtS21 promoter/GUS construct (pAN5) was introduced into both tobacco (by the leaf disc method, Horsch et al., 1985; Bogue et al. 1990 *Mol. Gen. Gen.*

221:49-57) and *Arabidopsis* Colombia ecotype via vacuum infiltration as described by Bechtold, et al. (1993) *C.R. Acad. Sci. Paris, 316:*1194-1199. Seeds were sterilized and selected on media containing 50 μg/ml kanamycin, 500 μg/ml carbenicillin.

GUS activity assay: Expression patterns of the reporter GUS gene were revealed by histochemical staining (Jefferson, et al., 1987a, *Plant Mol. Biol. Rep.* 5:387-405). Different tissues were stained in substrate solution containing 2 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid (X-Gluc) (Research Organics, Inc.), 0.5 mM potassium ferrocyanide, and 0.5 mM potassium ferricyanide in 50 mM sodium phosphate buffer, pH 7.0 at 37°C overnight, and then dehydrated successively in 20%, 40% and 80% ethanol (Jefferson, et al., 1987). Photographs were taken using an Axiophot (Zeiss) compound microscope or Olympus SZH10 dissecting microscope. Slides were converted to digital images using a Spring/Scan 35LE slide scanner (Polaroid) and compiled using Adobe Photoshop™ 3.0.5 and Canvas™ 3.5.

GUS activities were quantitatively measured by fluorometry using 2 mM 4-MUG (4-methylumbelliferyl-β-D-glucuronide) as substrate (Jefferson, et al., 1987). Developing *Arabidopsis* seeds were staged according to their colors, and other plant tissues were collected and kept at -80°C until use. Plant tissues were ground in extraction buffer containing 50 mM sodium phosphate, pH 7.0, 10 mM EDTA, 10 mM β-mercaptoethanol, 0.1% Triton X-100, and 0.1% sodium

lauryl sarcosine. The tissue debris was removed by 5 minutes centrifugation in a microfuge. The supernatant was aliquoted and mixed with substrate and incubated at 37°C for 1 hour. Three replicas were assayed for each sample. The reactions were stopped by adding 4 volumes of 0.2 M sodium carbonate. Fluorescence was read using a TKO-100 DNA fluorometer (Hoefer Scientific Instruments). Protein concentrations of the extracts were determined by the Bradford method (Bio Rad).

Expression patterns of AtS21 promoter/GUS in transgenic *Arabidopsis* and tobacco

In *Arabidopsis*, GUS activity was detected in green seeds, and node regions where siliques, cauline leaves and branches join the inflorescence stem (Figures 11A and 11B). No GUS activity was detected in any leaf, root, flower, silique coat, or the internode regions of the inflorescence stem. Detailed studies of the GUS expression in developing seeds revealed that the AtS21 promoter was only active in green seeds in which the embryos had already developed beyond heart stage (Figures 11C and 11G). The youngest embryos showing GUS activity that could be detected by histochemical staining were at early torpedo stage. Interestingly, the staining was only restricted to the lower part of the embryo including hypocotyl and embryonic radical. No staining was detected in the young cotyledons (Figures 11D and 11E). Cotyledons began to be stained when the embryos

were at late torpedo or even early cotyledon stage (Figure 11F and 11H). Later, the entire embryos were stained, and the staining became more intense as the embryos matured (Figures 11I and 11J). It was also observed that GUS gene expression was restricted to the embryos. Seed coat and young endosperm were not stained (Figure 11C).

GUS activity was also detected in developing seedlings. Young seedlings of 3-5 days old were stained everywhere. Although some root hairs close to the hypocotyl were stained (Figure 11K), most of the newly formed structures such as root hairs, lateral root primordia and shoot apex were not stained (Figures 11L and 11N). Later, the staining was restricted to cotyledons and hypocotyls when lateral roots grew from the elongating embryonic root. The staining on embryonic roots disappeared. No staining was observed on newly formed lateral roots, true leaves nor trichomes on true leaves (Figures 11M and 11N).

AtS21 promoter/GUS expression patterns in tobacco are basically the same as in *Arabidopsis*. GUS activity was only detected in late stage seeds and different node regions of mature plants. In germinating seeds, strong staining was detected throughout the entire embryos as soon as one hour after they were dissected from imbibed seeds. Mature endosperm, which *Arabidopsis* seeds do not have, but not seed coat was also stained (Figure 12A). The root tips of some young seedlings of one transgenic line

were not stained (Figure 12B). Otherwise, GUS expression patterns in developing tobacco seedlings were the same as in *Arabidopsis* seedlings (Figures 12B, 12C, and 12D). Newly formed structures such as lateral roots and true leaves were not stained.

AtS21 mRNA levels in developing seedlings

Since the observed strong activities of AtS21 promoter/GUS in both *Arabidopsis* and tobacco seedlings are not consistent with the seed-specific expression of oleosin genes, Northern analysis was carried out to determine if AtS21 mRNA was present in developing seedlings where the GUS activity was so strong. RNAs prepared from seedlings at different stages from 24 hours to 12 days were analyzed by Northern hybridization using AtS21 cDNA as the probe. Surprisingly, AtS21 mRNA was detected at a high level comparable to that in developing seeds in 24-48 hour imbibed seeds. The mRNA level dropped dramatically when young seedlings first emerged at 74 hours (Figures 13A and 13B). In 96 hour and older seedlings, no signal was detected even with a longer exposure (Figure 13B). The loadings of RNA samples were checked by hybridizing the same blot with a tubulin gene probe (Figure 13C) which was isolated and identified by EST analysis as described in Example 2. Since AtS21 mRNA was so abundant in seeds, residual AtS21 probes remained on the blot even after extensive stripping. These results indicated that AtS21 mRNA detected in imbibed seeds and very young seedlings are

the carry-over of AtS21 mRNA from dry seeds. It has recently been reported that an oleosin Atol2 mRNA (identical to AtS21) is most abundant in dry seeds (Kirik, et al., 1996 *Plant Mol. Biol. 31(2)*:413-417.) Similarly, the strong GUS activities in seedlings were most likely due to the carry-over of both β-glucuronidase protein and the *de novo* synthesis of β-glucuronidase from its mRNA carried over from the dry seed stage.

EXAMPLE 8

Activity comparison between the AtS21 promoter and the 35S promoter

The GUS activities in transgenic *Arabidopsis* developing seeds expressed by the AtS21 promoter were compared with those expressed by the 35S promoter in the construct pBI221 (Jefferson et al. *EMBO J. 6*:3901-3907). The seeds were staged according to their colors (Table 2). The earliest stage was from globular to late heart stage when the seeds were still white but large enough to be dissected from the siliques. AtS21 promoter activity was detected at a level about three times lower than that of the 35S promoter at this stage. 35S promoter activity remained at the same low level throughout the entire embryo development. In contrast, AtS21 promoter activity increased quickly as the embryos passed torpedo stage and reached the highest level of 25.25 pmole 4-MU/min. μg protein at mature stage (Figure 5-8). The peak activity of the AtS21 promoter is as much as 210 times higher than its lowest activity at globular to heart stage, and is close to 100 times higher than the 35S promoter activity at the same stage (Table 2). The activity levels of the AtS21 promoter are similar to those of another *Arabidopsis* oleosin promoter expressed in *Brassica napus* (Plant et al. 1994, *Plant mol. Biol. 25*:193-205. AtS21 promoter activity was also detected at background level in leaf. The high standard deviation, higher than the average itself, indicated that the GUS activity was

only detected in the leaves of some lines (Table 2). On the other hand, 35S promoter activity in leaf was more than 20 times higher than that in seed. The side by side comparisons of activities between AtS21 promoter and 35S promoter is shown in Figure 14.

Although the AtS21 promoter activity was about 3 times lower in dry seed of tobacco than in *Arabidopsis* dry seed, the absolute GUS activity was still higher than that expressed by the 35S promoter in *Arabidopsis* leaf (Table 2). No detectable AtS21 promoter activity was observed in tobacco leaf (Figure 14).

Comparison of the AtS21 promoter versus the 35S promoter revealed that the latter is not a good promoter to express genes at high levels in developing seeds. Because of its consistent low activities throughout the entire embryo development period, 35S promoter is useful for consistent low level expression of target genes. On the other hand, the AtS21 promoter is a very strong promoter that can be used to express genes starting from heart stage embryos and accumulating until the dry seed stage. The 35S promoter, although not efficient, is better than the AtS21 promoter in expressing genes in embryos prior to heart stage.

TABLE 2

GUS ACTIVITIES OF AtS21 and 35S PROMOTER/GUS CONSTRUCTS

COLOR STAGE	WHITE G-H	WHITE/YELLOW H-T	YELLOW T-C	LIGHT GREEN EARLY C	DARK GREEN LATE C	GREEN/YELLOW/BROWN MATURE	BROWN DRY SEED	LEA:
AtS21	0.12±0.17	1.35±1.57	6.77±1.25	18.99±3.75	21.85±4.45	25.25±4.64	24.38±10.85	0.08±0.
35S	0.30±0.06	0.25±0.08	0.29±0.04	0.28±0.03	0.33±0.06	0.26±0.04	0.31±0.02	6.56±0.
AtS21	(In tobacco)						8.81±0.21	0.01±0.

Abbreviations: G, globular stage; H, heart stage; T, torpedo stage; C, cotyledon stage. The GUS activities are in pmole 4-MU/μg protein.min. For AtS21 promoter the numbers are the average of five independent lines with standard deviations. Three repeats were assayed for each line. For 35S promoter the numbers are the average of three repeats of the same line with standard deviations.

EXAMPLE 9

Expression of the Borage Δ^6-Desaturase Gene Under the Control of the AtS21 Promoter and Comparison to Expression Under the Control of the CaMV 35S Promoter

In order to create an expression construct with the AtS21 promoter driving expression of the borage $\Delta6$-desaturase gene, the GUS coding fragment from pAN5 was removed by digestion with SmaI and EcoICR I. The cDNA insert of pAN1 (Example 2)was then excised by first digesting with XhoI (and filling in the residual overhang as above), and then digesting with SmaI. The resulting fragment was used to replace the excised portion of pAN5, yielding pAN3.

After transformation of tobacco and *Arabidopsis* following the methods of Example 7, levels of Δ^6-desaturase activity were monitored by assaying the corresponding fatty acid methyl esters of its reaction products, γ-linolenic acid (GLA) and octadecatetraenoic acid (OTA) using the methods referred to in Example 3. The GLA and OTA levels (Table 3) of the transgenic seeds ranged up to 6.7% of C18 fatty acids (Mean = 3.1%) and 2.8% (Mean = 1.1%), respectively. No GLA or OTA was detected in the leaves of these plants. In comparison, CaMV 35 S promoter/Δ^6-desaturase transgenic plants produced GLA levels in seeds ranging up to 3.1% of C18 fatty acids (Mean = 1.3%) and no measurable OTA in seeds.

TABLE 3

EXPRESSION OF THE BORAGE Δ^6-DESATURASE IN TRANSGENIC PLANTS

PROMOTER	PLANT	SEED				LEAF			
		GLA*	RANGE	OTA*	RANGE	GLA	RANGE	OTA	RANGE
Cauliflower mosaic virus 35S	tobacco	1.3	0.7-3.1	n.d		20	19-22	9.7	8-11
Arabidopsis oleosin	Arabidopsis	3.1	0-6.7	1.1	0-2.8	n.d.		n.d	

*mean value expressed as the percent of the C_{18} fatty acids
n.d. not detected

EXAMPLE 10

Transformation of Oilseed Rape With an Expression Cassette Which Comprises the Oleosin 5' Regulatory Region Linked to the Borage Delta 6-Desaturase Gene

Oilseed rape, Cv. Westar, was transformed with the strain of *Agrobacterium tumefaciens* EHA105 containing the plasmid pAN3 (i.e. the borage Δ6-desaturase gene under the control of the *Arabidopsis* oleosin promoter-Example 9).

Terminal internodes of Westar were co-cultivated for 2-3 days with induced *Agrobacterium tumefaciens* strain EHA105 (Alt-Moerbe et al. 1988 *Mol. Gen. Genet.* 213:1-8; James et al. 1993 *Plant Cell Reports* 12:559-563), then transferred onto regeneration medium (Boulter etal. 1990 *Plant Science* 70:91-99; Fry et al. 1987 *Plant Cell Reports* 6:321-325). The regenerated shoots were transferred to growth medium (Pelletier et al. 1983 *Mol.Gen. Menet.* 191:244-250), and a polymerase chain reaction (PCR) test was performed on leaf fragments to assess the presence of the gene.

DNA was isolated from the leaves according to the protocol of KM Haymes et al. (1996) *Plant Molecular Biology Reporter* 14(3):280-284, and resuspended in 100µl of water, without RNase treatment. 5µl of extract were used for the PCR reaction, in a final volume of 50µl. The reaction was performed in a Perkin-Elmer 9600 thermocycler, with the following cycles:

1 cycle: 95°C, 5 minutes
30 cycles: 95°C, 45 sec; 52°C, 45 sec
72°C, 1 minute

1 cycle: 72°C, 5 minutes

and the following primers (derived from near the metal box regions, as indicated in Fig. 1, SEQ. NO.:1):
5' TGG AAA TGG AAC CAT AA 3'
5' GGA AAC AAA TGA TGC TC 3'
Amplification of the DNA revealed the expected 549 base pair PCR fragment (Figure 17).

The positive shoots were transferred to elongation medium, then to rooting medium (DeBlock et al 1989 *Plant Physiol.* 91:694-701).Shoots with a well-developed root system were transferred to the greenhouse. When plants were well developed, leaves were collected for Southern analysis in order to assess gene copy number.

Genomic DNA was extracted according to the procedure of Bouchez et al. (1996) *Plant Molecular Biology Reporter* 14:115-123, digested with the restriction enzymes *Bgl* I and/or *Cla* I, electrophoretically separated on agarose gel (Maniatis et al. 1982, in *Molecular Cloning; a Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor/NY*), and prepared for transfer to nylon membranes (Nytran membrane, Schleicher & Schuell) according to the instructions of the manufacturer. DNA was then transferred to membranes overnight by capillary action using 20XSSC (Maniatis et al. 1982).

Following transfer, the membranes were crosslinked by UV (Stratagene) for 30 seconds and pre-hybridized for 1 hour at 65°C in 15 ml of a solution containing 6XSSC, 0.5%SDS and 2.25% w/w dehydrated skim milk in glass vials in hybridization oven (Appligene). The membranes were hybridized overnight in the same solution containing a denatured hybridization probe radiolabelled with ^{32}P to a specific activity of 10^8 cpm/µg by the random primer method (with the Ready-To-Go kit obtained from Pharmacia). The probe represents a PCR fragment of the borage delta 6-desaturase gene (obtained in the conditions and with the primers detailed above). After hybridization, the filters were washed at 65°C in 2XSSC, 0.1% SDS for 15 minutes, and 0.2XSSC, 0.1%SDS for 15 minutes. The membranes were then wrapped in Saran-Wrap and exposed to Kodak XAR film using an intensifying screen at -70°C in a light-proof cassette. Exposure time was generally 3 days.

The results obtained confirm the presence of the gene. According to the gene construct, the number of bands in each lane of DNA digested by *Bgl* I or *Cla* I represents the number of delta 6-desaturase genes present in the genomic DNA of the plant. The digestion with *Bgl* 1 and *Cla* 1 together generates a fragment of 3435 bp.

The term "comprises" or "comprising" is defined as specifying the presence of the stated features, integers, steps, or components as referred to in the claims, but does not preclude the presence or addition of one or more other features, integers, steps, components, or groups thereof.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Rhone Poulenc Agro
Thomas, Terry L.
Li, Zhongsen

(ii) TITLE OF INVENTION: AN OLEOSIN 5' REGULATORY REGION FOR THE MODIFICATION OF PLANT SEED LIPID COMPOSITION

(iii) NUMBER OF SEQUENCES: 35

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Scully, Scott, Murphy & Presser
(B) STREET: 400 Garden City Plaza
(C) CITY: Garden City
(D) STATE: New York
(E) COUNTRY: USA
(F) ZIP: 11530

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: 08/831,575
(B) FILING DATE: 9 April 1997
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: DiGiglio, Frank S.
(B) REGISTRATION NUMBER: 31,346
(C) REFERENCE/DOCKET NUMBER: 10203

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (516) 742-4343
(B) TELEFAX: (516) 742-4366

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1684 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 43..1387

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ATATCTGCCT ACCCTCCCAA AGAGAGTAGT CATTTTTCAT CA ATG GCT GCT CAA      54
                                               Met Ala Ala Gln
                                                1

ATC AAG AAA TAC ATT ACC TCA GAT GAA CTC AAG AAC CAC GAT AAA CCC    102
Ile Lys Lys Tyr Ile Thr Ser Asp Glu Leu Lys Asn His Asp Lys Pro
 5                  10                  15                  20

GGA GAT CTA TGG ATC TCG ATT CAA GGG AAA GCC TAT GAT GTT TCG GAT    150
Gly Asp Leu Trp Ile Ser Ile Gln Gly Lys Ala Tyr Asp Val Ser Asp
                25                  30                  35

TGG GTG AAA GAC CAT CCA GGT GGC AGC TTT CCC TTG AAG AGT CTT GCT    198
Trp Val Lys Asp His Pro Gly Gly Ser Phe Pro Leu Lys Ser Leu Ala
            40                  45                  50

GGT CAA GAG GTA ACT GAT GCA TTT GTT GCA TTC CAT CCT GCC TCT ACA    246
Gly Gln Glu Val Thr Asp Ala Phe Val Ala Phe His Pro Ala Ser Thr
        55                  60                  65

TGG AAG AAT CTT GAT AAG TTT TTC ACT GGG TAT TAT CTT AAA GAT TAC    294
Trp Lys Asn Leu Asp Lys Phe Phe Thr Gly Tyr Tyr Leu Lys Asp Tyr
    70                  75                  80

TCT GTT TCT GAG GTT TCT AAA GAT TAT AGG AAG CTT GTG TTT GAG TTT    342
Ser Val Ser Glu Val Ser Lys Asp Tyr Arg Lys Leu Val Phe Glu Phe
 85                  90                  95                 100

TCT AAA ATG GGT TTG TAT GAC AAA AAA GGT CAT ATT ATG TTT GCA ACT    390
Ser Lys Met Gly Leu Tyr Asp Lys Lys Gly His Ile Met Phe Ala Thr
                105                 110                 115

TTG TGC TTT ATA GCA ATG CTG TTT GCT ATG AGT GTT TAT GGG GTT TTG    438
Leu Cys Phe Ile Ala Met Leu Phe Ala Met Ser Val Tyr Gly Val Leu
            120                 125                 130

TTT TGT GAG GGT GTT TTG GTA CAT TTG TTT TCT GGG TGT TTG ATG GGG    486
Phe Cys Glu Gly Val Leu Val His Leu Phe Ser Gly Cys Leu Met Gly
        135                 140                 145

TTT CTT TGG ATT CAG AGT GGT TGG ATT GGA CAT GAT GCT GGG CAT TAT    534
Phe Leu Trp Ile Gln Ser Gly Trp Ile Gly His Asp Ala Gly His Tyr
    150                 155                 160

ATG GTA GTG TCT GAT TCA AGG CTT AAT AAG TTT ATG GGT ATT TTT GCT    582
Met Val Val Ser Asp Ser Arg Leu Asn Lys Phe Met Gly Ile Phe Ala
165                 170                 175                 180

GCA AAT TGT CTT TCA GGA ATA AGT ATT GGT TGG TGG AAA TGG AAC CAT    630
Ala Asn Cys Leu Ser Gly Ile Ser Ile Gly Trp Trp Lys Trp Asn His
                185                 190                 195
```

AAT GCA CAT CAC ATT GCC TGT AAT AGC CTT GAA TAT GAC CCT GAT TTA 678
Asn Ala His His Ile Ala Cys Asn Ser Leu Glu Tyr Asp Pro Asp Leu
200 205 210

CAA TAT ATA CCA TTC CTT GTT GTG TCT TCC AAG TTT TTT GGT TCA CTC 726
Gln Tyr Ile Pro Phe Leu Val Val Ser Ser Lys Phe Phe Gly Ser Leu
215 220 225

ACC TCT CAT TTC TAT GAG AAA AGG TTG ACT TTT GAC TCT TTA TCA AGA 774
Thr Ser His Phe Tyr Glu Lys Arg Leu Thr Phe Asp Ser Leu Ser Arg
230 235 240

TTC TTT GTA AGT TAT CAA CAT TGG ACA TTT TAC CCT ATT ATG TGT GCT 822
Phe Phe Val Ser Tyr Gln His Trp Thr Phe Tyr Pro Ile Met Cys Ala
245 250 255 260

GCT AGG CTC AAT ATG TAT GTA CAA TCT CTC ATA ATG TTG TTG ACC AAG 870
Ala Arg Leu Asn Met Tyr Val Gln Ser Leu Ile Met Leu Leu Thr Lys
265 270 275

AGA AAT GTG TCC TAT CGA GCT CAG GAA CTC TTG GGA TGC CTA GTG TTC 918
Arg Asn Val Ser Tyr Arg Ala Gln Glu Leu Leu Gly Cys Leu Val Phe
280 285 290

TCG ATT TGG TAC CCG TTG CTT GTT TCT TGT TTG CCT AAT TGG GGT GAA 966
Ser Ile Trp Tyr Pro Leu Leu Val Ser Cys Leu Pro Asn Trp Gly Glu
295 300 305

AGA ATT ATG TTT GTT ATT GCA AGT TTA TCA GTG ACT GGA ATG CAA CAA 1014
Arg Ile Met Phe Val Ile Ala Ser Leu Ser Val Thr Gly Met Gln Gln
310 315 320

GTT CAG TTC TCC TTG AAC CAC TTC TCT TCA AGT GTT TAT GTT GGA AAG 1062
Val Gln Phe Ser Leu Asn His Phe Ser Ser Ser Val Tyr Val Gly Lys
325 330 335 340

CCT AAA GGG AAT AAT TGG TTT GAG AAA CAA ACG GAT GGG ACA CTT GAC 1110
Pro Lys Gly Asn Asn Trp Phe Glu Lys Gln Thr Asp Gly Thr Leu Asp
345 350 355

ATT TCT TGT CCT CCT TGG ATG GAT TGG TTT CAT GGT GGA TTG CAA TTC 1158
Ile Ser Cys Pro Pro Trp Met Asp Trp Phe His Gly Gly Leu Gln Phe
360 365 370

CAA ATT GAG CAT CAT TTG TTT CCC AAG ATG CCT AGA TGC AAC CTT AGG 1206
Gln Ile Glu His His Leu Phe Pro Lys Met Pro Arg Cys Asn Leu Arg
375 380 385

AAA ATC TCG CCC TAC GTG ATC GAG TTA TGC AAG AAA CAT AAT TTG CCT 1254
Lys Ile Ser Pro Tyr Val Ile Glu Leu Cys Lys Lys His Asn Leu Pro
390 395 400

TAC AAT TAT GCA TCT TTC TCC AAG GCC AAT GAA ATG ACA CTC AGA ACA 1302
Tyr Asn Tyr Ala Ser Phe Ser Lys Ala Asn Glu Met Thr Leu Arg Thr
405 410 415 420

TTG AGG AAC ACA GCA TTG CAG GCT AGG GAT ATA ACC AAG CCG CTC CCG 1350
Leu Arg Asn Thr Ala Leu Gln Ala Arg Asp Ile Thr Lys Pro Leu Pro
425 430 435

AAG AAT TTG GTA TGG GAA GCT CTT CAC ACT CAT GGT T AAAATTACCC 1397
Lys Asn Leu Val Trp Glu Ala Leu His Thr His Gly
440 445

TTAGTTCATG TAATAATTTG AGATTATGTA TCTCCTATGT TTGTGTCTTG TCTTGGTTCT 1457

ACTTGTTGGA GTCATTGCAA CTTGTCTTTT ATGGTTTATT AGATGTTTTT TAATATATTT 1517

TAGAGGTTTT GCTTTCATCT CCATTATTGA TGAATAAGGA GTTGCATATT GTCAATTGTT 1577

GTGCTCAATA TCTGATATTT TGGAATGTAC TTTGTACCAC GTGGTTTTCA GTTGAAGCTC 1637

ATGTGTACTT CTATAGACTT TGTTTAAATG GTTATGTCAT GTTATTT 1684

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 135 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Ala Gln Ile Lys Lys Tyr Ile Thr Ser Asp Glu Leu Lys Asn
1 5 10 15

His Asp Lys Pro Gly Asp Leu Trp Ile Ser Ile Gln Gly Lys Ala Tyr
20 25 30

Asp Val Ser Asp Trp Val Lys Asp His Pro Gly Gly Ser Phe Pro Leu
35 40 45

Lys Ser Leu Ala Gly Gln Glu Val Thr Asp Ala Phe Val Ala Phe His
50 55 60

Pro Ala Ser Thr Trp Lys Asn Leu Asp Lys Phe Phe Thr Gly Tyr Tyr
65 70 75 80

Leu Lys Asp Tyr Ser Val Ser Glu Val Ser Lys Asp Tyr Arg Lys Leu
85 90 95

Val Phe Glu Phe Ser Lys Met Gly Leu Tyr Asp Lys Lys Gly His Ile
100 105 110

Met Phe Ala Thr Leu Cys Phe Ile Ala Met Leu Phe Ala Met Ser Val
115 120 125

Tyr Gly Val Leu Phe Cys Glu
130 135

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 834 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 31..603

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
TTAGCCTTTA CTCTATAGTT TTAGATAGAC ATG GCG AAT GTG GAT CGT GAT CGG       54
                                 Met Ala Asn Val Asp Arg Asp Arg
                                                 140

CGT GTG CAT GTA GAC CGT ACT GAC AAA CGT GTT CAT CAG CCA AAC TAC       102
Arg Val His Val Asp Arg Thr Asp Lys Arg Val His Gln Pro Asn Tyr
    145                 150                     155

GAA GAT GAT GTC GGT TTT GGT GGC TAT GGC GGT TAT GGT GCT GGT TCT       150
Glu Asp Asp Val Gly Phe Gly Gly Tyr Gly Gly Tyr Gly Ala Gly Ser
160                 165                 170                 175

GAT TAT AAG AGT CGC GGC CCC TCC ACT AAC CAA ATC TTG GCA CTT ATA       198
Asp Tyr Lys Ser Arg Gly Pro Ser Thr Asn Gln Ile Leu Ala Leu Ile
                180                 185                 190

GCA GGA GTT CCC ATT GGT GGC ACA CTG CTA ACC CTA GCT GGA CTC ACT       246
Ala Gly Val Pro Ile Gly Gly Thr Leu Leu Thr Leu Ala Gly Leu Thr
            195                 200                 205

CTA GCC GGT TCG GTG ATC GGC TTG CTA GTC TCC ATA CCC CTC TTC CTC       294
Leu Ala Gly Ser Val Ile Gly Leu Leu Val Ser Ile Pro Leu Phe Leu
        210                 215                 220

CTC TTC AGT CCG GTG ATA GTC CCG GCG GCT CTC ACT ATT GGG CTT GCT       342
Leu Phe Ser Pro Val Ile Val Pro Ala Ala Leu Thr Ile Gly Leu Ala
    225                 230                 235

GTG ACG GGA ATC TTG GCT TCT GGT TTG TTT GGG TTG ACG GGT CTG AGC       390
Val Thr Gly Ile Leu Ala Ser Gly Leu Phe Gly Leu Thr Gly Leu Ser
240                 245                 250                 255

TCG GTC TCG TGG GTC CTC AAC TAC CTC CGT GGG ACG AGT GAT ACA GTG       438
Ser Val Ser Trp Val Leu Asn Tyr Leu Arg Gly Thr Ser Asp Thr Val
                260                 265                 270
```

CCA GAG CAA TTG GAC TAC GCT AAA CGG CGT ATG GCT GAT GCG GTA GGC 486
Pro Glu Gln Leu Asp Tyr Ala Lys Arg Arg Met Ala Asp Ala Val Gly
275 280 285

TAT GCT GGT ATG AAG GGA AAA GAG ATG GGT CAG TAT GTG CAA GAT AAG 534
Tyr Ala Gly Met Lys Gly Lys Glu Met Gly Gln Tyr Val Gln Asp Lys
290 295 300

GCT CAT GAG GCT CGT GAG ACT GAG TTC ATG ACT GAG ACC CAT GAG CCG 582
Ala His Glu Ala Arg Glu Thr Glu Phe Met Thr Glu Thr His Glu Pro
305 310 315

GGT AAG GCC AGG AGA GGC TCA TAAGCTAATA TAAATTGCGG GAGTCAGTTG 633
Gly Lys Ala Arg Arg Gly Ser
320 325

GAAACGCGAT AAATGTAGTT TTACTTTTAT GTCCCAGTTT CTTTCCTCTT TTAAGAATAT 693

CTTTGTCTAT ATATGTGTTC GTTCGTTTTG TCTTGTCCAA ATAAAAATCC TTGTTAGTGA 753

AATAAGAAAT GAAATAAATA TGTTTTCTTT TTTGAGATAA CCAGAAATCT CATACTATTT 813

TCTAAAAAAA AAAAAAAAAA A 834

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 191 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ala Asn Val Asp Arg Asp Arg Arg Val His Val Asp Arg Thr Asp
1 5 10 15

Lys Arg Val His Gln Pro Asn Tyr Glu Asp Asp Val Gly Phe Gly Gly
20 25 30

Tyr Gly Gly Tyr Gly Ala Gly Ser Asp Tyr Lys Ser Arg Gly Pro Ser
35 40 45

Thr Asn Gln Ile Leu Ala Leu Ile Ala Gly Val Pro Ile Gly Gly Thr
50 55 60

Leu Leu Thr Leu Ala Gly Leu Thr Leu Ala Gly Ser Val Ile Gly Leu
65 70 75 80

Leu Val Ser Ile Pro Leu Phe Leu Leu Phe Ser Pro Val Ile Val Pro
85 90 95

Ala Ala Leu Thr Ile Gly Leu Ala Val Thr Gly Ile Leu Ala Ser Gly
100 105 110

Leu Phe Gly Leu Thr Gly Leu Ser Ser Val Ser Trp Val Leu Asn Tyr
115 120 125

Leu Arg Gly Thr Ser Asp Thr Val Pro Glu Gln Leu Asp Tyr Ala Lys
130 135 140

Arg Arg Met Ala Asp Ala Val Gly Tyr Ala Gly Met Lys Gly Lys Glu
145 150 155 160

Met Gly Gln Tyr Val Gln Asp Lys Ala His Glu Ala Arg Glu Thr Glu
165 170 175

Phe Met Thr Glu Thr His Glu Pro Gly Lys Ala Arg Arg Gly Ser
180 185 190

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 191 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Ala Asn Val Asp Arg Asp Arg Arg Val His Val Asp Arg Thr Asp
1 5 10 15

Lys Arg Val His Gln Pro Asn Tyr Glu Asp Asp Val Gly Phe Gly Gly
20 25 30

Thr Gly Gly Thr Gly Ala Gly Ser Asp Tyr Lys Ser Arg Gly Pro Ser
35 40 45

Thr Asn Gln Ile Leu Ala Leu Ile Ala Gly Val Pro Ile Gly Gly Thr
50 55 60

Leu Ile Thr Leu Ala Gly Leu Thr Leu Ala Gly Ser Val Ile Gly Leu
65 70 75 80

Leu Val Ser Ile Pro Leu Phe Leu Ile Phe Ser Pro Val Ile Val Pro
85 90 95

Ala Ala Leu Thr Ile Gly Leu Ala Val Thr Gly Ile Leu Ala Ser Gly
100 105 110

Leu Phe Gly Leu Thr Gly Leu Ser Ser Val Ser Trp Val Leu Asn Tyr
115 120 125

Leu Arg Gly Thr Ser Asp Thr Val Pro Glu Gln Leu Asp Tyr Ala Lys
130 135 140

Arg Arg Met Ala Asp Ala Val Gly Tyr Ala Gly Met Lys Gly Lys Glu
145 150 155 160

Met Gly Gln Tyr Val Gln Asp Lys Ala His Glu Ala Arg Glu Thr Glu
165 170 175

Phe Met Thr Glu Thr His Glu Pro Gly Lys Ala Arg Arg Gly Ser
180 185 190

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 191 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Ala Asn Val Asp Arg Asp Arg Arg Val His Val Asp Arg Thr Asp
1 5 10 15

Lys Arg Val His Gln Pro Asn Tyr Glu Asp Asp Val Gly Phe Gly Gly
20 25 30

Thr Gly Gly Thr Gly Ala Gly Ser Asp Tyr Lys Ser Arg Gly Pro Ser
35 40 45

Thr Asn Gln Ile Leu Ala Leu Ile Ala Gly Val Pro Ile Gly Gly Thr
50 55 60

Leu Ile Thr Leu Ala Gly Leu Thr Leu Ala Gly Ser Val Ile Gly Leu
65 70 75 80

Leu Val Ser Ile Pro Leu Phe Leu Ile Phe Ser Pro Val Ile Val Pro
85 90 95

Ala Ala Leu Thr Ile Gly Leu Ala Val Thr Gly Ile Leu Ala Ser Gly
100 105 110

Leu Phe Gly Leu Thr Gly Leu Ser Ser Val Ser Trp Val Leu Asn Tyr
115 120 125

Leu Arg Gly Thr Ser Asp Thr Val Pro Glu Gln Leu Asp Tyr Ala Lys
130 135 140

Arg Arg Met Ala Asp Ala Val Gly Tyr Ala Gly Met Lys Gly Lys Glu
145 150 155 160

Met Gly Gln Tyr Val Gln Asp Lys Ala His Glu Ala Arg Glu Thr Glu
165 170 175

Phe Met Thr Glu Thr His Glu Pro Gly Lys Ala Arg Arg Gly Pro
180 185 190

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 78 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Phe Gly Leu Thr Gly Leu Ser Ser Val Ser Trp Val Leu Gln Leu Pro
1 5 10 15

Pro Trp Ala Ser Asp Thr Val Pro Glu Gln Val Asp Tyr Ala Lys Arg
20 25 30

Arg Met Ala Asp Ala Val Gly Tyr Ala Gly Met Lys Gly Lys Glu Met
35 40 45

Gly Gln Tyr Val Gln Asp Lys Ala His Glu Ala Arg Glu Thr Glu Phe
50 55 60

Met Thr Glu Thr His Glu Pro Gly Lys Ala Arg Arg Gly Ser
65 70 75

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 173 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Ala Asp Thr Ala Arg Gly Thr His His Asp Ile Ile Gly Arg Asp
1 5 10 15

Gln Tyr Pro Met Met Gly Arg Asp Arg Asp Gln Tyr Gln Met Ser Gly
20 25 30

Arg Gly Ser Asp Tyr Ser Lys Ser Arg Gln Ile Ala Lys Ala Ala Thr
35 40 45

Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Ser Leu Thr Leu
50 55 60

Val Gly Thr Val Leu Ala Leu Thr Val Ala Thr Pro Leu Leu Val Leu
65 70 75 80

Phe Ser Pro Ile Leu Val Pro Ala Leu Ile Thr Val Ala Leu Leu Ile
85 90 95

Thr Gly Phe Leu Ser Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Val
100 105 110

Phe Ser Trp Ile Tyr Lys Tyr Ala Thr Gly Glu His Pro Gln Gly Ser
115 120 125

Asp Lys Leu Asp Ser Ala Arg Met Lys Leu Gly Ser Lys Ala Gln Asp
130 135 140

Leu Lys Asp Arg Ala Gln Tyr Tyr Gly Gln Gln His Thr Gly Gly Glu
145 150 155 160

His Asp Arg Asp Arg Thr Arg Gly Gly Gln His Thr Thr
165 170

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 141 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Ala Asp Gln Thr Arg Thr His His Glu Met Ile Ser Arg Asp Ser
1 5 10 15

Thr Gln Glu Ala His Pro Lys Ala Arg Gln Trp Val Lys Ala Ala Thr
20 25 30

Ala Val Thr Ala Gly Gly Ser Leu Leu Val Leu Ser Gln Leu Thr Leu
35 40 45

Ala Gly Thr Val Ile Ala Leu Thr Val Ala Thr Pro Leu Leu Val Ile
50 55 60

Phe Ser Pro Val Leu Val Pro Ala Val Val Thr Val Ala Leu Ile Ile
65 70 75 80

Thr Gly Phe Leu Ala Ser Gly Gly Phe Gly Ile Ala Ala Ile Thr Ala
85 90 95

Phe Ser Trp Leu Tyr Arg His Trp Thr Gly Ser Gly Ser Asp Lys Ile
100 105 110

Glu Trp Ala Arg Met Lys Val Gly Ser Arg Val Gln Asp Thr Lys Tyr
115 120 125

Gly Gln His Trp Ile Gly Val Gln His Gln Gln Val Ser
130 135 140

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 199 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Ala Asp Thr His Arg Val Asp Arg Thr Asp Arg His Phe Gln Phe
1 5 10 15

Gln Ser Pro Tyr Glu Gly Gly Arg Gly Gln Gly Gln Tyr Glu Gly Asp
20 25 30

Arg Gly Tyr Gly Gly Gly Gly Tyr Lys Ser Met Met Pro Glu Ser Gly
35 40 45

Pro Ser Ser Thr Gln Val Leu Ser Leu Leu Ile Gly Val Pro Val Val
50 55 60

Gly Ser Leu Ile Ala Leu Ala Gly Leu Leu Leu Ala Gly Ser Val Ile
65 70 75 80

Gly Leu Met Val Ala Leu Pro Leu Phe Leu Ile Phe Ser Pro Val Ile
85 90 95

Val Pro Ala Gly Leu Thr Ile Gly Leu Ala Met Thr Gly Phe Leu Ala
100 105 110

Ser Gly Met Phe Gly Leu Thr Gly Leu Ser Ser Ile Ser Trp Val Met
115 120 125

Asn Tyr Leu Arg Gly Thr Ala Arg Thr Val Pro Glu Gln Leu Glu Tyr
130 135 140

Ala Lys Arg Arg Met Ala Asp Ala Val Gly Tyr Ala Gly Gln Lys Gly
145 150 155 160

Lys Glu Met Gly Gln His Val Gln Asn Lys Ala Gln Asp Val Lys Gln
165 170 175

Tyr Asp Ile Ser Lys Pro His Asp Thr Thr Thr Lys Gly His Glu Thr
180 185 190

Gln Gly Gly Thr Thr Ala Ala
195

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 199 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Ala Asp Thr His Arg Val Asp Arg Thr Asp Arg His Phe Gln Phe
1 5 10 15

Gln Ser Pro Tyr Glu Gly Gly Arg Gly Gln Gly Gln Tyr Glu Gly Asp
20 25 30

Arg Gly Tyr Gly Gly Gly Gly Tyr Lys Ser Met Met Pro Glu Ser Gly
35 40 45

Pro Ser Ser Thr Gln Val Leu Ser Leu Leu Ile Gly Val Pro Val Val
50 55 60

Gly Ser Leu Ile Ala Leu Ala Gly Leu Leu Ile Ala Gly Ser Val Ile
65 70 75 80

Gly Leu Met Val Ala Leu Pro Leu Phe Leu Ile Phe Ser Pro Val Ile
85 90 95

Val Pro Ala Ala Leu Thr Ile Gly Leu Ala Met Thr Gly Phe Leu Ala
100 105 110

Ser Gly Met Phe Gly Leu Thr Gly Leu Ser Ser Ile Ser Trp Val Met
115 120 125

Asn Tyr Leu Arg Gly Thr Arg Arg Thr Val Pro Glu Gln Leu Glu Tyr
130 135 140

Ala Lys Arg Arg Met Ala Asp Ala Val Gly Tyr Ala Gly Gln Lys Gly
145 150 155 160

Lys Glu Met Gly Gln His Val Gln Asn Lys Ala Gln Asp Val Lys Gln
165 170 175

Tyr Asp Ile Ser Lys Pro His Asp Thr Thr Thr Lys Gly His Glu Thr
180 185 190

Gln Gly Arg Thr Thr Ala Ala
195

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1267 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAGCTCGATC ACACAAAGAA AACGTCAAAT GGATCATACT GGGCCCATTT TGCAGACCAA 60
GAGAAAGTGA GAGAGAGTTG TCCTCTCGTT ATCAAGTAAC AGTAGACCAC CACTAAACCG 120
CCAATAGCTT ATAATCAAAA TAGAAAGGTC TAATAACAGA AACAAATGAA AAAGCCTTGT 180
TCCATGGACT GCCTACCCGA ATTGATTGAT TCGACTAGTT TTTCTTCTTC TTTGATTAAG 240
ACCTCCGTAA GAAAAATGGT ACTACTAAAG CCACTCGCTA CCAAAACTAA ACCATTCCAG 300
ACTGTAACTG GACCAATATT TCTAAACTGT AACCAGATCT CAAACATATA AACTAATTAA 360
GAACTATAAC CATTAACCGT AAAAATAAAT TTACTACAGT AAAAAATTAT ACTAATTTCA 420
GCTATGATGG AATTTCAGCT CTTAAGAGTT GTGGAAATCA AGTAAACCTA AAATCCTAAT 480
AATATTCTTC ATCCTTATTT TTGTTTCACA TGCATGCTGT CCAATCTGTT ATTAGCATTT 540
GAAAGCCTAA AATTCTATAT ACAGTACAAT AAATCTAATT AATTTTCATT ACTAATAAAA 600
TGCTTCATAT ATACTCTTGT ATTTATAAAT CATCCGTTAT CGTTACTATA CCTTTATACA 660
TCATCCTACA TTCATACCTA AGCTAGCAAA GCAAACTACT AAAAGGGTCG TCAACGCAAG 720
TTATTTGCTA GTTGGTGCAT ACTACACACG GCTACGGCAA CATTAAGTAA CACATTAAGA 780
GGTGTTTTCT TAATGTAGTA TGGTAATTAT ATTTATTTCA AAACTTGGAT TAGATATAAA 840
GGTACAGGTA GATGAAAAAT ATTTGGTTAG CGGGTTGAGA TTAAGCGGAT ATAGGAGGCA 900
TATATACAGC TGTGAGAAGA AGAGGGATAA ATACAAAAAG GGAAGGATGT TTTTGCCGAC 960
AGAGAAAGGT AGATTAAGTA GGCATCGAGA GGAGAGCAAT TGTAAAATGG ATGATTTGTT 1020
TGGTTTTGTA CGGTGGAGAG AAGAACGAAA AGATGATCAG GTAAAAAATG AAACTTGGAA 1080
ATCATGCAAA GCCACACCTC TCCCTTCAAC ACAGTCTTAC GTGTCGTCTT CTCTTCACTC 1140
CATATCTCCT TTTTATTACC AAGAAATATA TGTCAATCCC ATTTATATGT ACGTTCTCTT 1200

AGACTTATCT CTATATACCC CCTTTTAATT TGTGTGCTCT TAGCCTTTAC TCTATAGTTT 1260

TAGATAG 1267

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTAATACGAC TCACTATAGG GC 22

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GGGGATCCTA TACTAAAACT ATAGAGTAAA GG 32

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Trp Ile Gly His Asp Ala Gly His
1 5

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Asn Val Gly His Asp Ala Asn His
1 5

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Val Leu Gly His Asp Cys Gly His
1 5

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Val Ile Ala His Glu Cys Gly His
1 5

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Val Ile Gly His Asp Cys Ala His
1 5

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Val Val Gly His Asp Cys Gly His
1 5

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

His Asn Ala His His
1 5

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

His Asn Tyr Leu His His
1 5

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

His Arg Thr His His
1 5

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

His Arg Arg His His
1 5

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

His Asp Arg His His
1 5

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

His Asp Gln His His
1 5

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

His Asp His His His
1 5

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

His Asn His His His
1 5

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Phe Gln Ile Glu His His
1 5

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

His Gln Val Thr His His
1 5

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

His Val Ile His His
1 5

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

His Val Ala His His
1 5

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

His Ile Pro His His
1 5

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

His Val Pro His His
1 5

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1941 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GAGCTCGATC ACACAAAGAA AACGTCAAAT GGATCATACT GGGCCCATTT TGCAGACCAA 60

GAGAAAGTGA GAGAGAGTTG TCCTCTCGTT ATCAAGTAAC AGTAGACCAC CACTAAACCG 120

CCAATAGCTT ATAATCAAAA TAGAAAGGTC TAATAACAGA AACAAATGAA AAAGCCTTGT 180

TCCATGGACT GCCTACCCGA ATTGATTGAT TCGACTAGTT TTTCTTCTTC TTTGATTAAG 240

ACCTCCGTAA GAAAAATGGT ACTACTAAAG CCACTCGCTA CCAAAACTAA ACCATTCCAG 300

ACTGTAACTG GACCAATATT TCTAAACTGT AACCAGATCT CAAACATATA AACTAATTAA 360

GAACTATAAC CATTAACCGT AAAAATAAAT TTACTACAGT AAAAAATTAT ACTAATTTCA 420

GCTATGATGG AATTTCAGCT CTTAAGAGTT GTGGAAATCA AGTAAACCTA AAATCCTAAT 480

AATATTCTTC ATCCTTATTT TTGTTTCACA TGCATGCTGT CCAATCTGTT ATTAGCATTT 540

GAAAGCCTAA AATTCTATAT ACAGTACAAT AAATCTAATT AATTTTCATT ACTAATAAAA 600

TGCTTCATAT ATACTCTTGT ATTTATAAAT CATCCGTTAT CGTTACTATA CCTTTATACA 660
TCATCCTACA TTCATACCTA AGCTAGCAAA GCAAACTACT AAAAGGGTCG TCAACGCAAG 720
TTATTTGCTA GTTGGTGCAT ACTACACACG GCTACGGCAA CATTAAGTAA CACATTAAGA 780
GGTGTTTTCT TAATGTAGTA TGGTAATTAT ATTTATTTCA AAACTTGGAT TAGATATAAA 840
GGTACAGGTA GATGAAAAAT ATTTGGTTAG CGGGTTGAGA TTAAGCGGAT ATAGGAGGCA 900
TATATACAGC TGTGAGAAGA AGAGGGATAA ATACAAAAAG GGAAGGATGT TTTTGCCGAC 960
AGAGAAAGGT AGATTAAGTA GGCATCGAGA GGAGAGCAAT TGTAAAATGG ATGATTTGTT 1020
TGGTTTTGTA CGGTGGAGAG AAGAACGAAA AGATGATCAG GTAAAAAATG AAACTTGGAA 1080
ATCATGCAAA GCCACACCTC TCCCTTCAAC ACAGTCTTAC GTGTCGTCTT CTCTTCACTC 1140
CATATCTCCT TTTTATTACC AAGAAATATA TGTCAATCCC ATTTATATGT ACGTTCTCTT 1200
AGACTTATCT CTATATACCC CCTTTTAATT TGTGTGCTCT TAGCCTTTAC TCTATAGTTT 1260
TAGATAGACA TGGCGAATGT GGATCGTGAT CGGCGTGTGC ATGTAGACCG TACTGACAAA 1320
CGTGTTCATC AGCCAAACTA CGAAGATGAT GTCGGTTTTG GTGGCTATGG CGGTTATGGT 1380
GCTGGTTCTG ATTATAAGAG TCGCGGCCCC TCCACTAACC AAGTATTTTT GTGGTCTCTT 1440
TAGTTTTTCT TGTGTTTTCC TATGATCACG CTCTCCAAAC TATTTGAAGA TTTTCTGTAA 1500
ATTCATTTTA AACAGAAAGA TAAATAAAAT AGTGAAGAAC CATAGGAATC GTACGTTACG 1560
TTAATTATTT CCTTTTAGTT CTTAAGTCCT AATTAGGATT CCTTTAAAAG TTGCAACAAT 1620
CTAATTGTTC ACAAAATGAG TAAAGTTTGA AACAGATTTT TATACACCAC TTGCATATGT 1680
TTATCATGGT GATGCATGCT TGTTAGATAA ACTCGATATA ATCAATACAT GCAGATCTTG 1740
GCACTTATAG CAGGAGTCCA TTGGTGGCAC ACTGCTAACC CTAGCTGGAC TCACTCTAGC 1800
CGGTTCGGTG ATCGGCTTGC TAGTCTCCAT ACCCCTCTTC CTCCTCTTCA GTCCGGTGAT 1860
AGTCCCGGCG GCTCTCACTA TTGGGCTTGC TGTGACGGGA ATCTTGGCTT CTGGTTTGTT 1920
TGGGTTGACG GGTCTGAGCT C 1941

What is claimed is:

1. An isolated nucleic acid encoding an oleosin 5' regulatory region which directs seed-specific expression selected from the groups consisting of the nucleotide sequence set forth in SEQ ID NO:12, the nucleotide sequence set forth in SEQ ID NO:12 having an insertion, deletion, or substitution of one or more nucleotides, or a contiguous fragment of the nucleotide sequence set forth in SEQ ID NO:12.

2. An expression cassette which comprises the oleosin 5' regulatory region of Claim 1 operably linked to at least one of a nucleic acid encoding a heterologous gene or a nucleic acid encoding a sequence complementary to a native plant gene.

3. The expression cassette of Claim 2 wherein the heterologous gene is at least one of a fatty acid synthesis gene or a lipid metabolism gene.

4. The expression cassette of Claim 3 wherein the heterologous gene is selected from the group consisting of an acetyl-coA carboxylase gene, a ketoacyl synthase gene, a malonyl transacylase gene, a lipid desaturase gene, an acyl carrier protein (ACP) gene, a thioesterase gene, an acetyl transacylase gene, or an elongase gene.

5. The expression cassette of Claim 4 wherein the lipid desaturase gene is selected from the group consisting of a $\Delta6$-desaturase gene, a $\Delta12$-desaturase gene, and a $\Delta15$-desaturase gene.

6. An expression vector which comprises the expression cassette of any one of Claims 2-5.

7. A cell comprising the expression cassette of any one of Claims 2-5.

8. A cell comprising the expression vector of Claim 6.

9. The cell of Claim 7 wherein said cell is a bacterial cell or a plant cell.

10. The cell of Claim 8 wherein said cell is a bacterial cell or a plant cell.

11. A transgenic plant comprising the expression cassette of any one of Claims 2-5.

12. A transgenic plant comprising the expression vector of Claim 6.

13. A plant which has been regenerated from the plant cell of Claim 9.

14. A plant which has been regenerated from the plant cell of Claim 10.

15. The plant of Claim 12 or 13 wherein said plant is at least one of a sunflower, soybean, maize, cotton, tobacco, peanut, oil seed rape or *Arabidopisis* plant.

16. Progeny of the plant of Claim 11 or 12.

17. Seed from the plant of Claim 11 or 12.

18. A method of producing a plant with increased levels of a product of a fatty acid synthesis gene or a lipid metabolism gene which comprises:

(a) transforming a plant cell with an expression vector comprising the isolated nucleic acid of Claim 1 operably linked to at least one of an

isolated nucleic acid coding for a fatty acid synthesis gene or a lipid metabolism gene; and

(b) regenerating a plant with increased levels of the product of said fatty acid synthesis or said lipid metabolism gene from said plant cell.

19. A method of producing a plant with increased levels of gamma linolenic acid (GLA) content which comprises:

(a) transforming a plant cell with an expression vector comprising the isolated nucleic acid of Claim 1 operably linked to a Δ6-desaturase gene; and

(b) regenerating a plant with increased levels of GLA from said plant cell.

20. The method of Claim 19 wherein said Δ6-desaturase gene is at least one of a cyanobacterial Δ6-desaturase gene or a Borage Δ6-desaturase gene.

21. The method of any one of Claims 18-20 wherein said plant is a sunflower, soybean, maize, tobacco, cotton, peanut, oil seed rape or *Arabidopsis* plant.

22. The method of Claim 18 wherein said fatty acid synthesis gene or said lipid metabolism gene is at least one of a lipid desaturase, an acyl carrier protein (ACP) gene, a thioesterase gene an elongase gene, an acetyl transacylase gene, an acetyl-coA carboxylase gene, a ketoacyl synthase gene, or a malonyl transacylase gene.

23. A method of inducing production of at least one of gamma linolenic acid (GLA) or

octadecatetraeonic acid (OTA) in a plant deficient or lacking in GLA which comprises transforming said plant with an expression vector comprising an the isolated nucleic acid of Claim 1 operably linked to a Δ6-desaturase gene and regenerating a plant with increased levels of at least one of GLA or OTA.

24. A method of decreasing production of a fatty acid synthesis or lipid metabolism gene in a plant which comprises:

(a) transforming a plant cell with an expression vector comprising the isolated nucleic acid of Claim 1 operably linked to a nucleic acid sequence complementary to a fatty acid synthesis or lipid metabolism gene; and

(b) regenerating a plant with decreased production of said fatty acid synthesis or said lipid metabolism gene.

25. A method of cosuppressing a native fatty acid synthesis or lipid metabolism gene in a plant which comprises:

(a) transforming a cell of the plant with an expression vector comprising the isolated nucleic acid of Claim 1 operably linked to a nucleic acid sequence encoding a fatty acid synthesis or lipid metabolism gene native to the plant; and

(b) regenerating a plant with decreased production of said fatty acid synthesis or said lipid metabolism gene.

FIG.1A

```
2
ata tct gcc tac cct ccc aaa gag agt agt cat ttt tca tca atg gct gct caa atc aag
                                                         M   A   A   Q   I   K

62
aaa tac att acc tca gat gaa ctc aag aac cac gat aaa ccc gga gat cta tgg atc tcg
 K   Y   I   T   S   D   E   L   K   N   H   D   K   P   G   D   L   W   I   S

122
att caa ggg aaa gcc tat gat gtt tcg gat tgg gtg aaa gac cat cca ggt ggc agc ttt
 I   Q   G   K   A   Y   D   V   S   D   W   V   K   D   H   P   G   G   S   F

182
ccc ttg aag agt ctt gct ggt caa gag gta act gat gca ttt gtt gca ttc cat cct gcc
 P   L   K   S   L   A   G   Q   E   V   T   D   A   F   V   A   F   H   P   A

242
tct aca tgg aag aat ctt gat aag ttt ttc act ggg tat tat ctt aaa gat tac tct gtt
 S   T   W   K   N   L   D   K   F   F   T   G   Y   Y   L   K   D   Y   S   V

302
tct gag gtt tct aaa gat tat agg aag ctt gtg ttt gag ttt tct aaa atg ggt ttg tat
 S   E   V   S   K   D   Y   R   K   L   V   F   E   F   S   K   M   G   L   Y

362
gac aaa aaa ggt cat att atg ttt gca act ttg tgc ttt ata gca atg ctg ttt gct atg
 D   K   K   G   H   I   M   F   A   T   L   C   F   I   A   M   L   F   A   M
```

FIG.1B

```
422
agt gtt tat ggg gtt ttg ttt tgt gag ggt gtt ttg gta cat ttg ttt tct ggg tgt ttg
S   V   Y   G   V   L   F   C   E   G   V   L   V   H   L   F   S   G   C   L

482
atg ggg ttt ctt tgg att cag agt ggt tgg att gga cat gat gct ggg cat tat atg gta
M   G   F   L   W   I   Q   S   G   W   I   G   H   D   A   G   H   Y   M   V

542
gtg tct gat tca agg ctt aat aag ttt atg ggt att ttt gct gca aat tgt ctt tca gga
V   S   D   S   R   L   N   K   F   M   G   I   F   A   A   N   C   L   S   G

602
ata agt att ggt tgg tgg aaa tgg aac cat aat gca cat cac att gcc tgt aat agc ctt
I   S   I   G   W   W   K   W   N   H   N   A   H   H   I   A   C   N   S   L

662
gaa tat gac cct gat tta caa tat ata cca ttc ctt gtt gtg tct tcc aag ttt ttt ggt
E   Y   D   P   D   L   Q   Y   I   P   F   L   V   V   S   S   K   F   F   G

722
tca ctc acc tct cat ttc tat gag aaa agg ttg act ttt gac tct tta tca aga ttc ttt
S   L   T   S   H   F   Y   E   K   R   L   T   F   D   S   L   S   R   F   F
```

FIG.1C

782
gta agt tat caa cat tgg aca ttt tac cct att atg tgt gct gct agg ctc aat atg tat
V S Y Q H W T F Y P I M C A A R L N M Y

842
gta caa tct ctc ata atg ttg ttg acc aag aga aat gtg tcc tat cga gct cag gaa ctc
V Q S L I M L L T K R N V S Y R A Q E L

902
ttg gga tgc cta gtg ttc tcg att tgg tac ccg ttg ctt gtt tct tgt ttg cct aat tgg
L G C L V F S I W Y P L L V S C L P N W

962
ggt gaa aga att atg ttt gtt att gca agt tta tca gtg act gga atg caa caa gtt cag
G E R I M F V I A S L S V T G M Q Q V Q

1022
ttc tcc ttg aac cac ttc tct tca agt gtt tat gtt gga aag cct aaa ggg aat aat tgg
F S L N H F S S S V Y V G K P K G N N W

1082
ttt gag aaa caa acg gat ggg aca ctt gac att tct tgt cct cct tgg atg gat tgg ttt
F E K Q T D G T L D I S C P P W M D W F

FIG.1D

```
1142
cat ggt gga ttg caa ttc caa att gag cat cat ttg ttt ccc aag atg cct aga tgc aac
 H   G   G   L   Q   F   Q   I   E   H   H   L   F   P   K   M   P   R   C   N

1202
ctt agg aaa atc tcg ccc tac gtg atc gag tta tgc aag aaa cat aat ttg cct tac aat
 L   R   K   I   S   P   Y   V   I   E   L   C   K   K   H   N   L   P   Y   N

1262
tat gca tct ttc tcc aag gcc aat gaa atg aca ctc aga aca ttg agg aac aca gca ttg
 Y   A   S   F   S   K   A   N   E   M   T   L   R   T   L   R   N   T   A   L

1322
cag gct agg gat ata acc aag ccg ctc ccg aag aat ttg gta tgg gaa gct ctt cac act
 Q   A   R   D   I   T   K   P   L   P   K   N   L   V   W   E   A   L   H   T

1382
cat ggt taa aat tac cct tag ttc atg taa taa ttt gag att atg tat ctc cta tgt ttg
 H   G   *

1442
tgt ctt gtc ttg gtt cta ctt gtt gga gtc att gca act tgt ctt tta tgg ttt att aga

1502
tgt ttt tta ata tat ttt aga ggt ttt gct ttc atc tcc att att gat gaa taa gga gtt
```

FIG.1E

1562
gca tat tgt caa ttg ttg tgc tca ata tct gat att ttg gaa tgt act ttg tac cac tgt

1622
gtt ttc agt tga agc tca tgt gta ctt cta tag act ttg ttt aaa tgg tta tgt cat gtt

1682
att t

FIG. 2

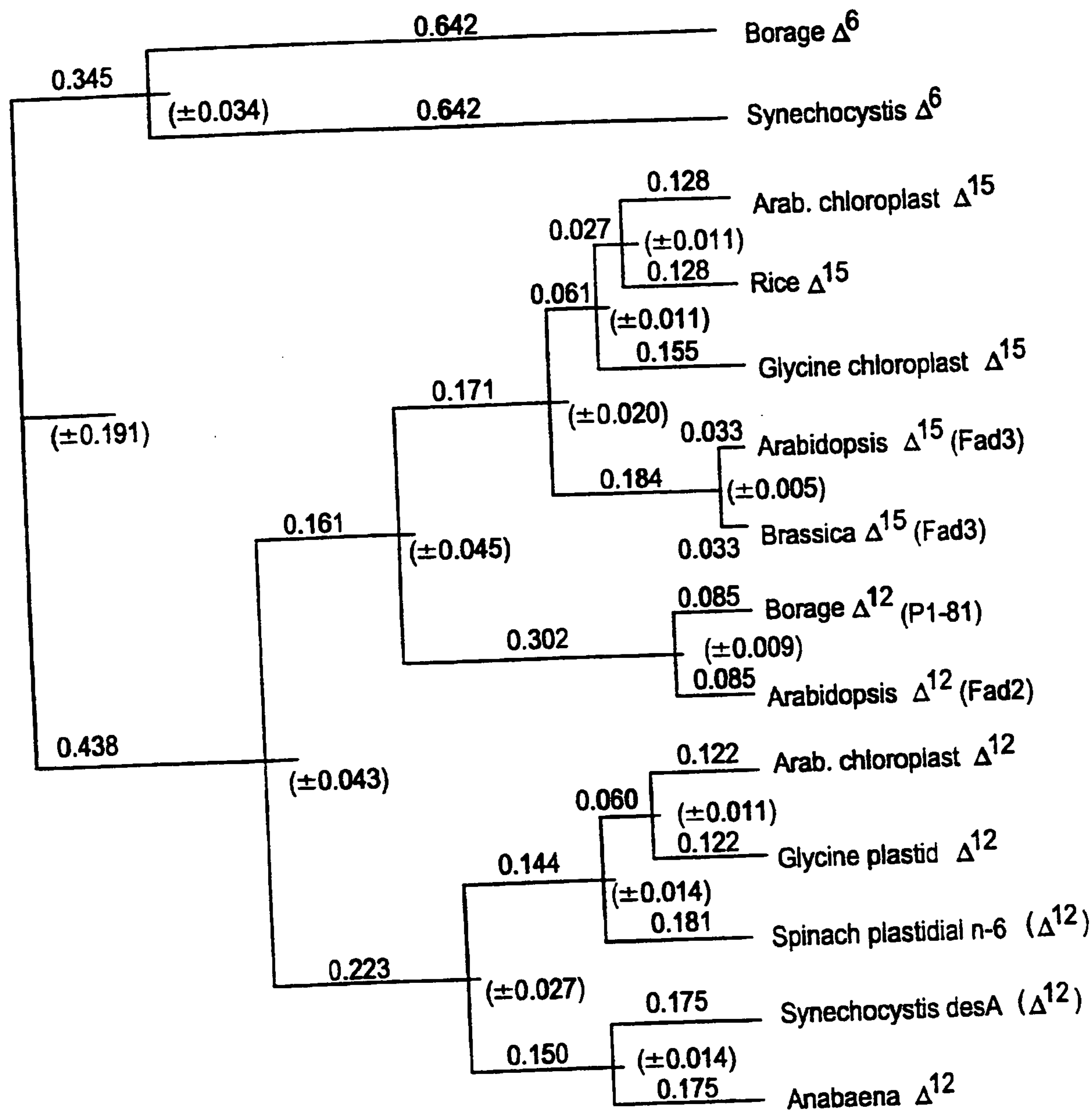

0.642
Borage Δ6
0.345
(±0.034)
0.642
Synechocystis Δ6
0.128
Arab. chloroplast Δ15
0.027
(±0.011)
0.128
Rice Δ15
0.061
(±0.011)
0.155
Glycine chloroplast Δ15
0.171
(±0.191)
(±0.020)
0.033
Arabidopsis Δ15 (Fad3)
0.184
(±0.005)
0.161
(±0.045)
0.033
Brassica Δ15 (Fad3)
0.085
Borage Δ12 (P1-81)
0.302
(±0.009)
0.085
Arabidopsis Δ12 (Fad2)
0.438
(±0.043)
0.122
Arab. chloroplast Δ12
0.060
(±0.011)
0.144
0.122
Glycine plastid Δ12
(±0.014)
0.181
Spinach plastidial n-6 (Δ12)
0.223
(±0.027)
0.175
Synechocystis desA (Δ12)
0.150
(±0.014)
0.175
Anabaena Δ12

FIG. 3A

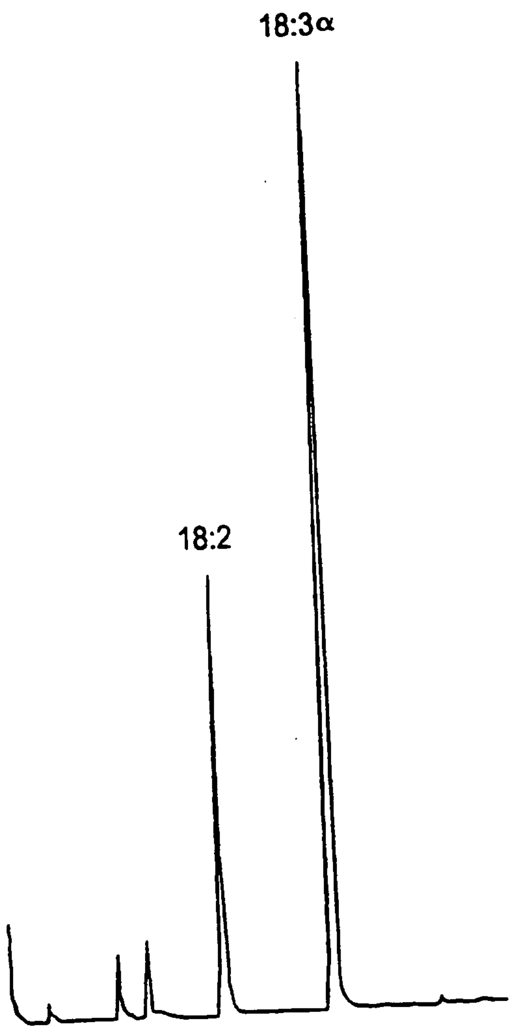


FIG. 3B

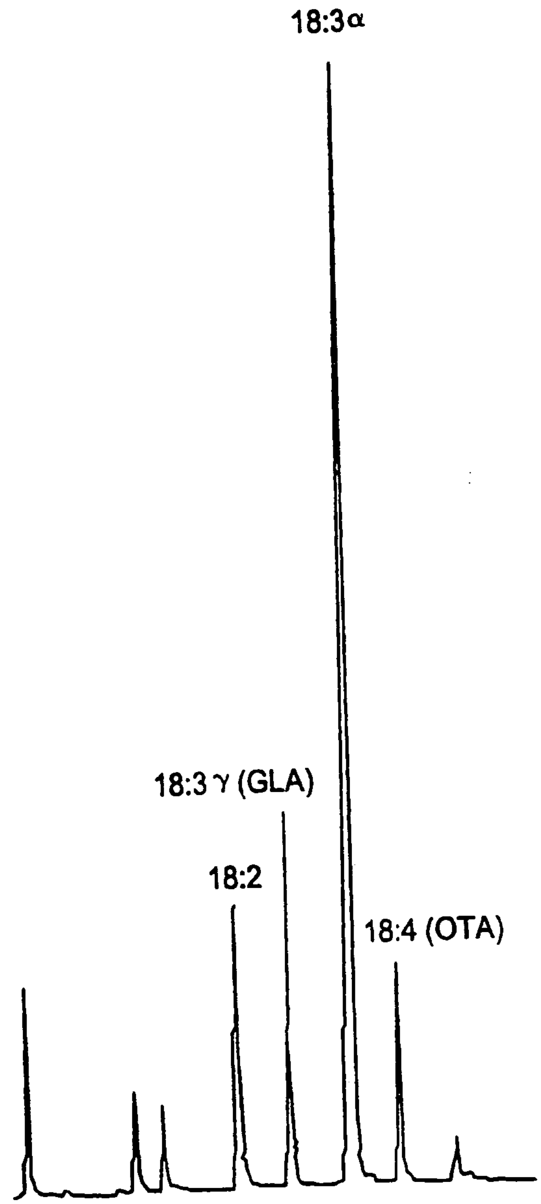


FIG.4

```
1  tta gcc ttt act cta tag ttt tag ata gac atg gcg aat gtg gat cgt gat cgg cgt gtg
1                                          M   A   N   V   D   R   D   R   R   V
61 cat gta gac cgt act gac aaa cgt gtt cat cag cca aac tac gaa gat gat gtc ggt ttt
11 H   V   D   R   T   D   K   R   V   H   Q   P   N   Y   E   D   D   V   G   F
121ggt ggc tat ggc ggt tat ggt gct ggt tct gat tat aag agt cgc ggc ccc tcc act aac
31 G   G   Y   G   G   Y   G   A   G   S   D   Y   K   S   R   G   P   S   T   N
181caa atc ttg gca ctt ata gca gga gtt ccc att ggt ggc aca ctg cta acc cta gct gga
51 Q   I   L   A   L   I   A   G   V   P   I   G   G   T   L   L   T   L   A   G
241ctc act cta gcc ggt tcg gtg atc ggc ttg cta gtc tcc ata ccc ctc ttc ctc ctc ttc
71 L   T   L   A   G   S   V   I   G   L   L   V   S   I   P   L   F   L   L   F
301agt ccg gtg ata gtc ccg gcg gct ctc act att ggg ctt gct gtg acg gga atc ttg gct
91 S   P   V   I   V   P   A   A   L   T   I   G   L   A   V   T   G   I   L   A
361tct ggt ttg ttt ggg ttg acg ggt ctg agc tcg gtc tcg tgg gtc ctc aac tac ctc cgt
111S   G   L   F   G   L   T   G   L   S   S   V   S   W   V   L   N   Y   L   R
421ggg acg agt gat aca gtg cca gag caa ttg gac tac gct aaa cgg cgt atg gct gat gcg
131G   T   S   D   T   V   P   E   Q   L   D   Y   A   K   R   R   M   A   D   A
481gta ggc tat gct ggt atg aag gga aaa gag atg ggt cag tat gtg caa gat aag gct cat
151V   G   Y   A   G   M   K   G   K   E   M   G   Q   Y   V   Q   D   K   A   H
541gag gct cgt gag act gag ttc atg act gag acc cat gag ccg ggt aag gcc agg aga ggc
171E   A   R   E   T   E   F   M   T   E   T   H   E   P   G   K   A   R   R   G
601tca taa gct aat ata aat tgc ggg agt cag ttg gaa acg cga taa atg tag ttt tac ttt
191S   *
661tat gtc cca gtt tct ttc ctc ttt taa gaa tat ctt tgt cta tat atg tgt tcg ttc gtt
721ttg tct tgt cca aat aaa aat cct tgt tag tga aat aag aaa tga aat aaa tat gtt ttc
781ttt ttt gag ata acc aga aat ctc ata cta ttt tct aaa aaa aaa aaa aaa aaa
```

FIG.5

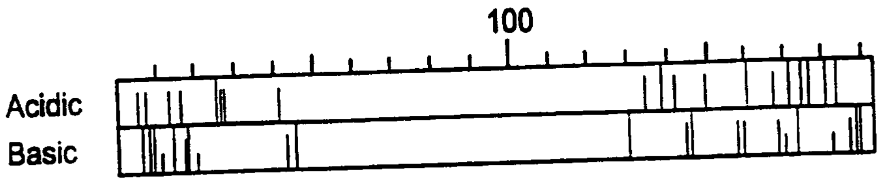


FIG.6

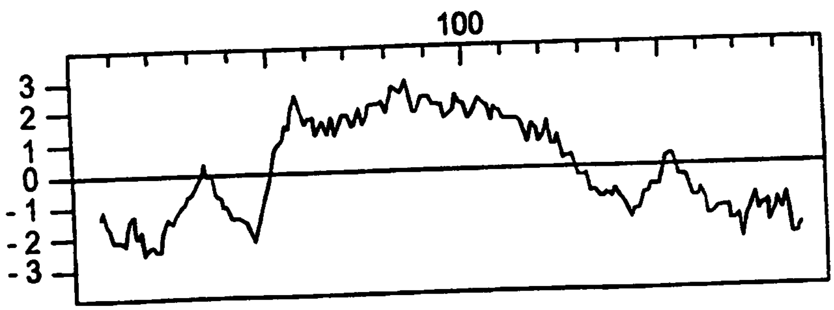


FIG.7A

```
AtS21   MANVDRDRRV HVDRTDKRVH -QPNYEDDVG ---F-G--GY GGYGAGSDYK 43
X91918  MANVDRDRRV HVDRTDKRVH -QPNYEDDVG ---F-G--GY GGYGAGSDYK 43
Z29859  ---------- ---------- ---------- ---------- ----------
X62352  MADTARG--- -THHDIIGRD QYPMMGRDRD QYQKSGRG-- ------SDY- 37
Atol3   MADQTR---- -THHE----- ---MISRDST Q--------- ------EAH- 21
X91956  MADTHR---- -VDRTDRHFQ FQSPYEGGRG QGQYEGDRGY GGGGYKSMMP 45
L40954  MADTHR---- -VDRTDRHFQ FQSPYEGGRG QGQYEGDRGY GGGGYKSMMP 45

AtS21   SRGPSTNQIL ALIAGVPIGG TLITLAGLTL AGSVIGLIVS IPIFLLFSPV 93
X91918  SRGPSTNQIL ALIAGVPIGG TLITLAGLTL AGSVIGLIVS IPIFLLFSPV 93
Z29859  ---------- ---------- ---------- ---------- ----------
X62352  ---SKSRQIA KAATAVTAGG SLIVLSSLTL VGTVIALTVA TPLLVIFSPI 84
Atol3   ---PKARQMV KAATAVTAGG SLIVLSQLTL AGTVIALTVA TPLLVIFSPV 68
X91956  ESGPSSTQVL SLLIGVPVVG SLIALAGLLL AGSVIGLMVA LPLFLIFSPV 95
L40954  ESGPSSTQVL SLLIGVPVVG SLIALAGLLL AGSVIGLMVA LPLFLIFSPV 95

AtS21   IVPAALTIGL AVTGILASGL FGLTGLSSVS WVLNYLRGTS DTVPEQLDYA 143
X91918  IVPAALTIGL AVTGILASGL FGLTGLSSVS WVLNYLRGTS DTVPEQLDYA 143
Z29859  ---------- ---------- FGLTGLSSVS WVLQLPPWAS DTVPEQVDYA 30
X62352  IVPALITVAL LITGFLSSGG FGIAAITVFS WIYKYATGEH PQGSDKLDSA 134
Atol3   IVPAVVTVAL LITGFLASGG FGIAAITAFS WLYRHHTGS- --GSDKIENA 115
X91956  IVPAGITIGL AMTGFLASGM FGLTGLSSIS WVMNYLRGTK RTVPEQLEYA 145
L40954  IVPAAITIGL AMTGFLASGM FGLTGLSSIS WVMNYLRGTR RTVPEQLEYA 145
```

FIG.7B

```
AtS21     KRRMADAVGY AGMKGKEMGQ YVQDKAHEAR ETEF------ MTETHEPGKA 187
X91918    KRRMADAVGY AGMKGKEMGQ YVQDKAHEAR ETEF------ MTETHEPGKA 187
Z29859    KRRMADAVGY AGMKGKEMGQ YVQDKAHEAR ETEF------ MTETHEPGKA 74
X62352    RMKLGSKA-- QDLKDRA-QY YGQQHTGGEH DRDRTRGGQH TT-------- 173
Atol3     RMKVGSRV-- QDTK------ YGQHNIGVQH QQ-------- VS-------- 141
X91956    KRRMADAVGY AGQKGKEMGQ HVQNKAQDVK QYDISKPHDT TTKGHETQGG 195
L40954    KRRMADAVGY AGQKGKEMGQ HVQNKAQDVK QYDISKPHDT TTKGHETQGR 195

AtS21     RRGS = Z54164/Atol2                                  191
X91918    RRGP                                                 191
Z29859    RRGS                                                 78
X62352    ----                                                 173
Atol3     ----                                                 141
X91956    TTAA                                                 199
L40954    TTAA                                                 199
```

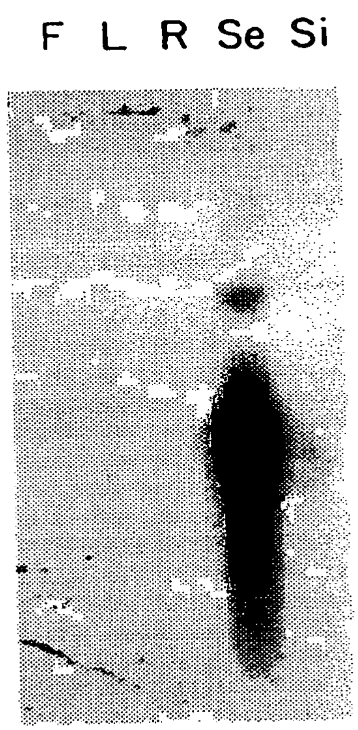


FIG.8A

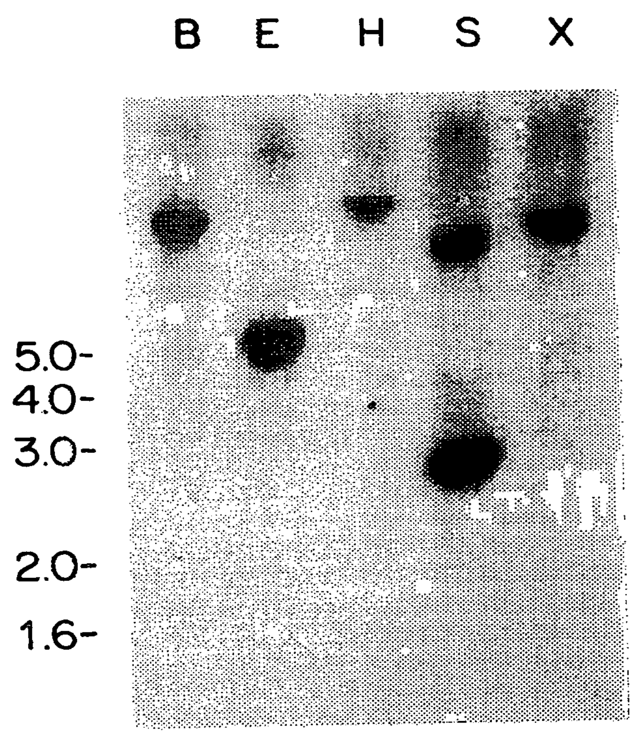


FIG.8B

FIG.9

1 GAGCTCGATCACACAAAGAAAACGTCAAATGGATCATACTGGGCCCATTTTGCAGACCAAGAGAAAGTGAGAGAGAGTTG
81 TCCTCTCGTTATCAAGTAACAGTAGACCACCACTAAACCGCCAATAGCTTATAATCAAAATAGAAAGGTCTAATAACAGA
161 AACAAATGAAAAAGCCTTGTTCCATGGACTGCCTACCCGAATTGATTGATTCGACTAGTTTTTCTTCTTCTTTGATTAAG
241 ACCTCCGTAAGAAAAATGGTACTACTAAAGCCACTCGCTACCAAAACTAAACCATTCCAGACTGTAACTGGACCAATATT
321 TCTAAACTGTAACCAGATCTCAAACATATAAACTAATTAAGAACTATAACCATTAACCGTAAAAATAAATTTACTACAGT
401 AAAAAATTATACTAATTTCAGCTATGATGGAATTTCAGCTCTTAAGAGTTGTGGAAATCAAGTAAACCTAAAATCCTAAT
481 AATATTCTTCATCCTTATTTTTGTTTCACATGCATGCTGTCCAATCTGTTATTAGCATTTGAAAGCCTAAAATTCTATAT
561 ACAGTACAATAAATCTAATTAATTTTCATTACTAATAAAATGCTTCATATATACTCTTGTATTTATAAATCATCCGTTAT
641 CGTTACTATACCTTTATACATCATCCTACATTCATACCTAAGCTAGCAAAGCAAACTACTAAAAGGGTCGTCAACGCAAG
721 TTATTTGCTAGTTGGTGCATACTACACACGGCTACGGCAACATTAAGTAACACATTAAGAGGTGTTTTCTTAATGTAGTA
801 TGGTAATTATATTTATTTCAAAACTTGGATTAGATATAAAGGTACAGGTAGATGAAAAATATTTGGTTAGCGGGTTGAGA
881 TTAAGCGGATATAGGAGGCATATATACAGCTGTGAGAAGAAGAGGGATAAATACAAAAAGGGAAGGATGTTTTTGCCGAC
961 AGAGAAAGGTAGATTAAGTAGGCATCGAGAGGAGAGCAATTGTAAAATGGATGATTTGTTTGGTTTTGTACGGTGGAGAG
1041 AAGAACGAAAAGATGATCAGGTAAAAAATGAAACTTGGAAATCATGCAAAGCCACACCTCTCCCTTCAACACAGTCTTAC
1121 GTGTCGTCTTCTCTTCACTCCATATCTCCTTTTTATTACCAAGAAATATATGTCAATCCCATTTATATGTACGTTCTCTT
1201 AGACTTATCTCTATATACCCCCTTTTAATTTGTGTGCTCttag$\boxed{\texttt{cctttactctatagttttagatag}}$acatggcgaatgt
1281 ggatcgtgatcggcgtgtgcatgtagaccgtactgacaaacgtgttcatcagccaaactacgaagatgatgtcggttttg
1361 gtggctatggcggttatggtgctggttctgattataagagtcgcggcccctccactaaccaaGTATTTTTGTGGTCTCTT
1441 TAGTTTTTCTTGTGTTTTCCTATGATCACGCTCTCCAAACTATTTGAAGATTTTCTGTAAATTCATTTTAAACAGAAAGA
1521 TAAATAAAATAGTGAAGAACCATAGGAATCGTACGTTACGTTAATTATTTCCTTTTAGTTCTTAAGTCCTAATTAGGATT
1601 CCTTTAAAAGTTGCAACAATCTAATTGTTCACAAAATGAGTAAAGTTTGAAACAGATTTTTATACACCACTTGCATATGT
1681 TTATCATGGTGATGCATGCTTGTTAGATAAACTCGATATAATCAATACATGCAGatcttggcacttatagcaggagttcc
1761 cattggtggcacactgctaaccctagctggactcactctagccggttcggtgatcggcttgctagtctccatacccctct
1821 tcctcctcttcagtccggtgatagtcccggcggctctcactattgggcttgctgtgacgggaatcttggcttctggtttg
1901 tttgggttgacgggtctgagctc

FIG.10

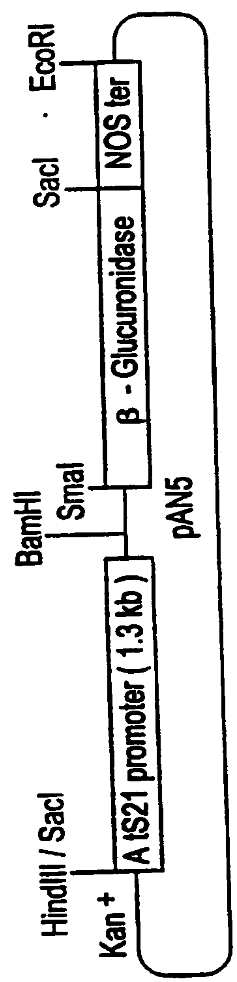

EcoRI
SacI
NOS ter
β - Glucuronidase
BamHI
SmaI
pAN5
A tS21 promoter (1.3 kb)
HindIII / SacI
Kan +

FIG. 11A

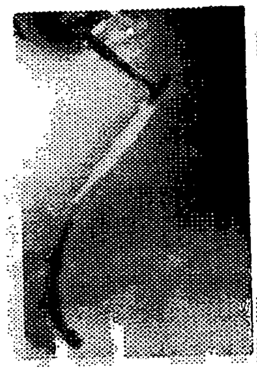


FIG. 11B

FIG. 11C

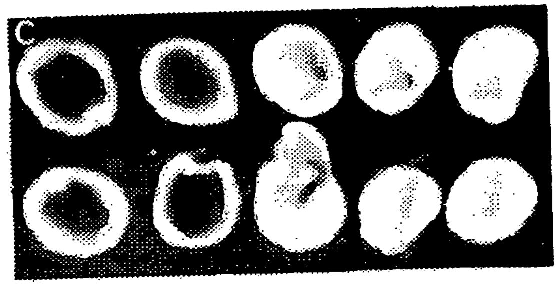


FIG. 11D

FIG. 11E

FIG. 11F

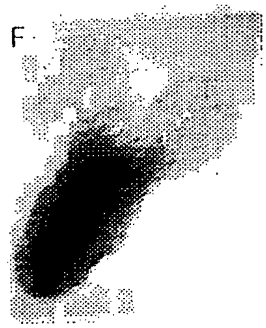


FIG. 11G

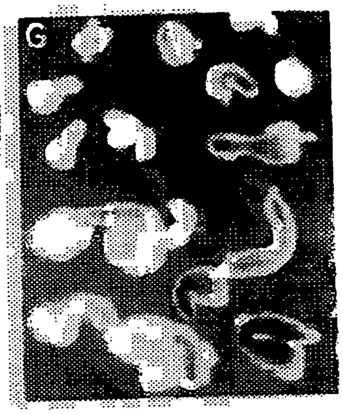


FIG. 11H

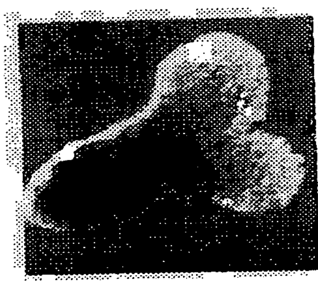


FIG. 11I

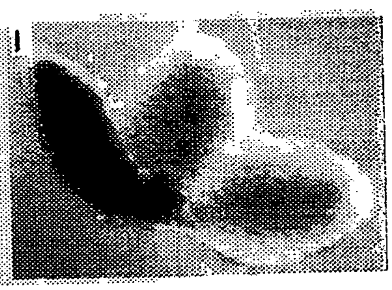


FIG. 11J

FIG. 11K

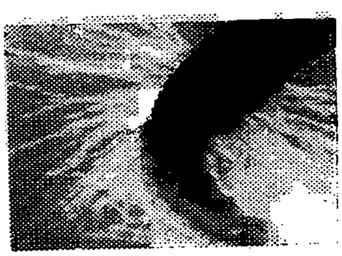


FIG. 11L

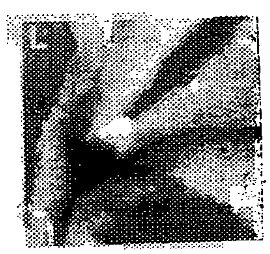


FIG. 11M

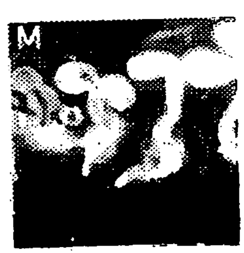


FIG. 11N

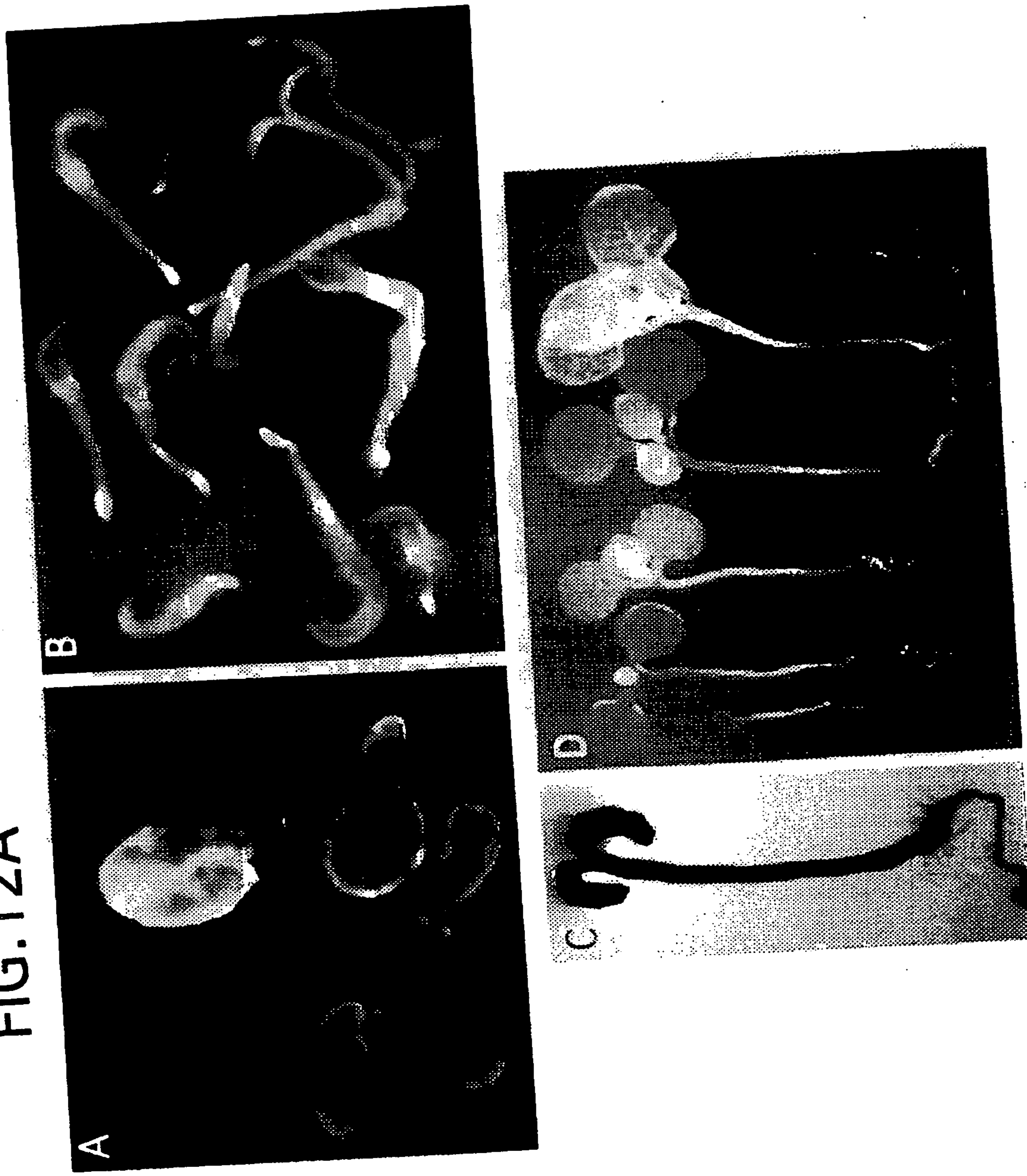


FIG.12A

FIG.12B

FIG.12C

FIG.12D

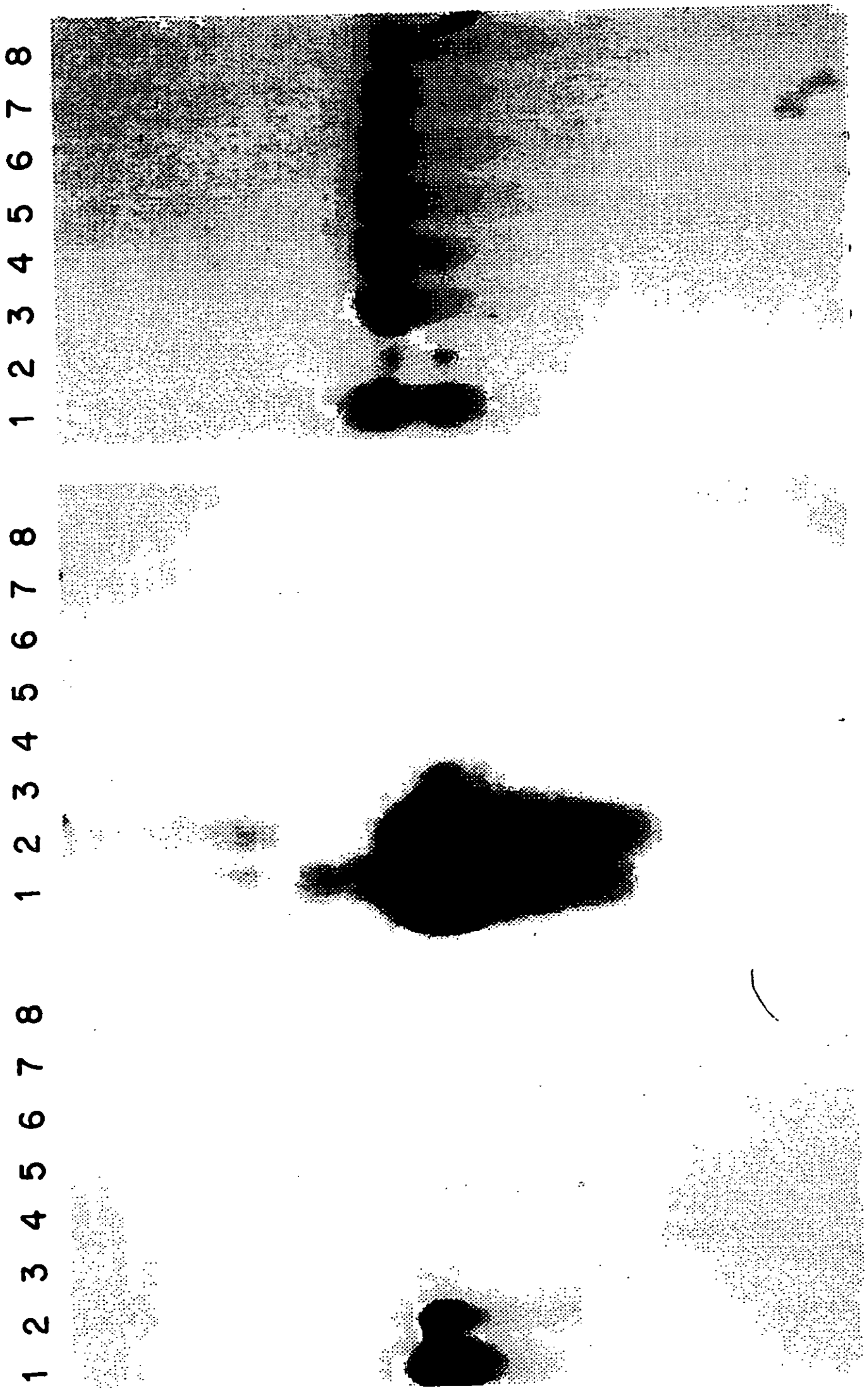
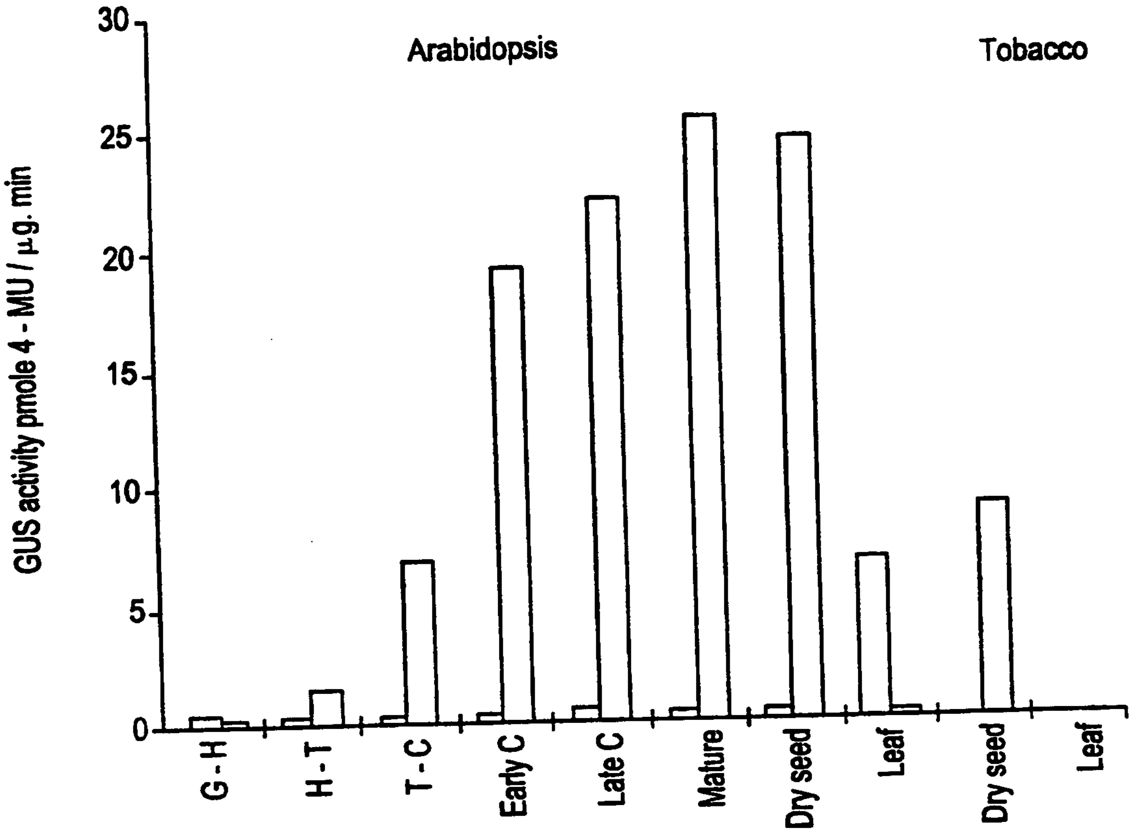


FIG.13A

FIG.13B

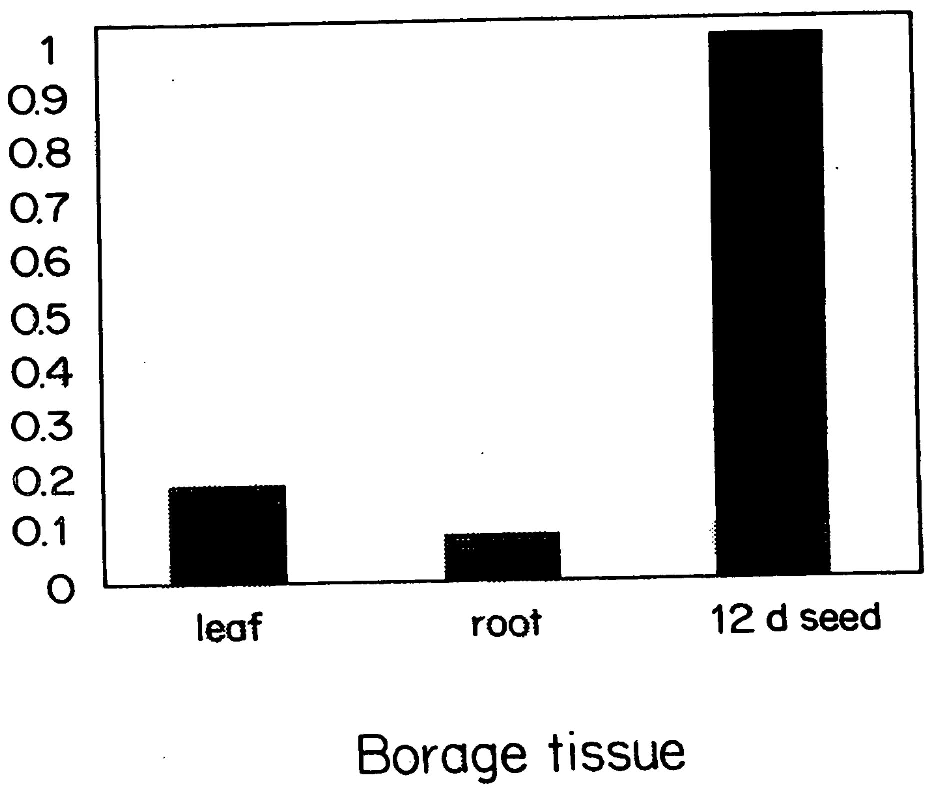
FIG.13C

FIG.14



30
25
20
15
10
5
0
GUS activity pmole 4 - MU / µg. min
Arabidopsis
Tobacco
G - H
H - T
T - C
Early C
Late C
Mature
Dry seed
Leaf
Dry seed
Leaf

FIG. 15B



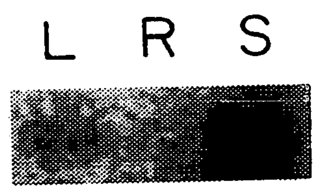


FIG. 15A

FIG.16A

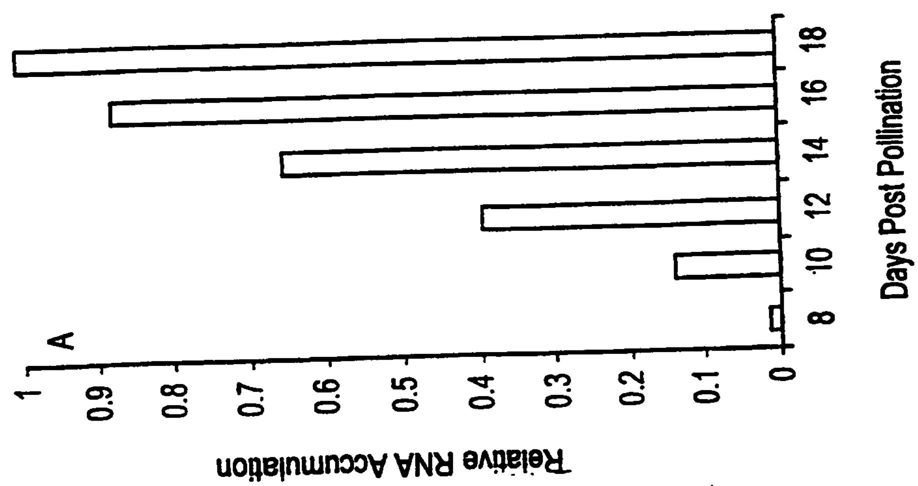


FIG.16B

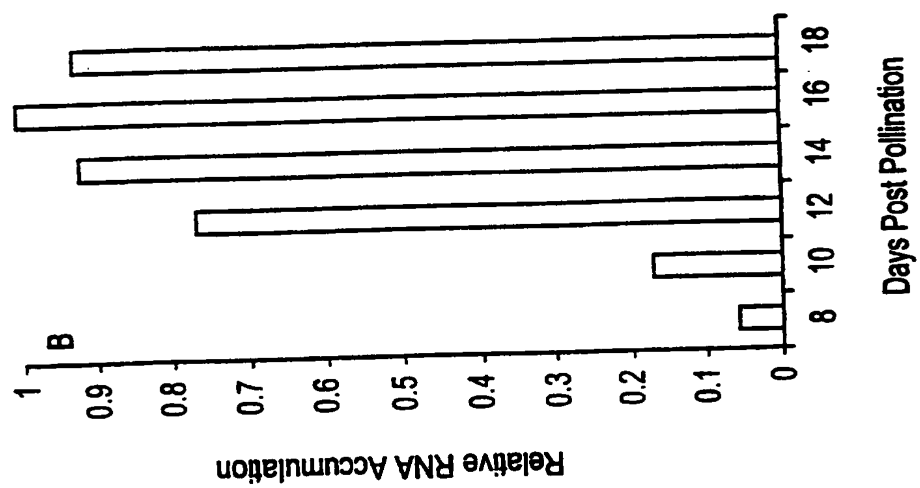


FIG.16C

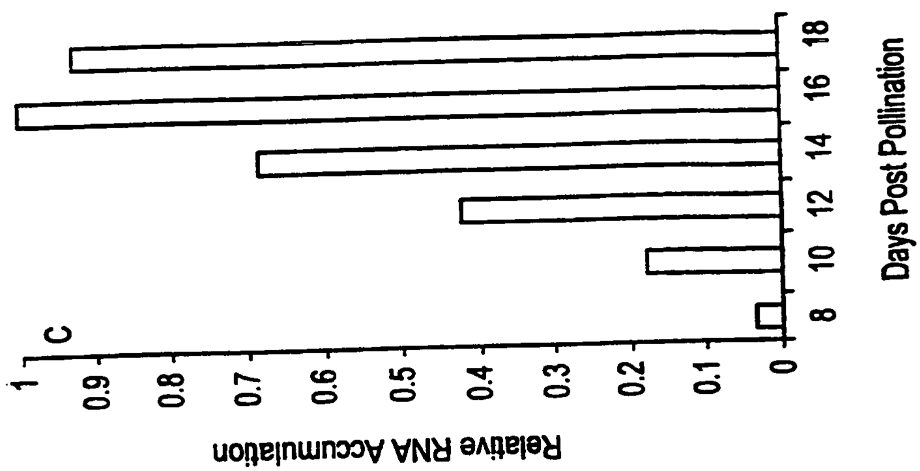


FIGURE 17

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/US 98/07179

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/82 C12N15/29 C12N15/53 A01H5/00 A01H5/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	"ISOLATION AND CHARACTERIZATION OF ARABIDOPSIS EMBRYO-SPECIFIC GENES (VIRTUAL SUBTRACTION, DNA BINDING SITES, GENE ISOLATION)" LI, ZHONGSEN 'PH.D.!; THOMAS, TERRY L. 'ADVISER! TEXAS A&M UNIVERSITY, May 1997, XP002075900 see page 107 - page 128 ---	1-17
A	KIRIK, V., ET AL.: "Two new oleosin isoforms with altered expression patterns in seeds of the Arabidopsis mutant fus3" PLANT MOLECULAR BIOLOGY, vol. 31, 1996, pages 413-417, XP002075901 see figure 1 ---	1

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 September 1998	21/09/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Maddox, A

2

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 93 20216 A (MOLONEY MAURICE M ;UNIV TECHNOLOGIES INT (CA)) 14 October 1993 see the whole document	1,2
A	PLANT, A.L.., ET AL.: "Regulation of an Arabidopsis oleosin gene promoter in transgenic Brassica napus" PLANT MOLECULAR BIOLOGY, vol. 25, 1994, pages 193-205, XP002075902 cited in the application see the whole document	1,2
A	WO 96 23892 A (CALGENE INC ;DEHESH KATAYOON (US); VOELKER TONI ALOIS (US); HAWKIN) 8 August 1996 see page 23, line 13 - page 26, line 5 see page 32, line 28 - page 33, line 2	1-4
A	WO 96 21022 A (RHONE POULENC AGROCHIMIE) 11 July 1996 see the whole document	1-23
A	BEREMAND, P.D., ET AL.: "Production of gamma-linolenic acid by transgenic plants expressing cyanobacterial or plant delta6-desaturase genes" PHYSIOL., BIOCHEM. MOL. BIOL. PLANT LIPIDS, (PROC. INT. SYMP. PLANT LIPIDS), 12TH, 1997, pages 351-353, XP002076486 see page 353	1-23
O,A	& BEREMAND, ET AL.: DISCLOSURE AT THE12TH INTERNATIONAL SYMPOSIUM ON PLANT LIPIDS, HELD JULY 7-12, 1996, UNIVERSITY OF TORONTO, TORONTO, 1996,	1-23
A	WO 93 11245 A (DU PONT) 10 June 1993 see the whole document	5,24,25
A	WO 94 18337 A (MONSANTO CO ;UNIV MICHIGAN (US); GIBSON SUSAN IRMA (US); KISHORE G) 18 August 1994 see the whole document	5,24,25
A	WO 94 11516 A (DU PONT ;LIGHTNER JONATHAN EDWARD (US); OKULEY JOHN JOSEPH (US)) 26 May 1994 see the whole document	5,24,25
A	WO 96 06936 A (DU PONT ;HITZ WILLIAM DEAN (US)) 7 March 1996 see page 48, line 23 - page 53, line 10	5,24,25

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	TOEPFER R ET AL: "MODIFICATION OF PLANT LIPID SYNTHESIS" SCIENCE, vol. 268, 5 May 1995, pages 681-685, XP002014017 see the whole document ---	1-25
A	GIBSON, S., ET AL.: "Cloning of a temperature-regulated gene encoding a chloroplast omega-3 desaturase from Arabidopsis thaliana" PLANT PHYSIOLOGY, vol. 106, 1994, pages 1615-1621, XP002075189 see the whole document -----	5,25

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 98/07179

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9320216 A	14-10-1993	WO 9321320 A	28-10-1993
		US 5650554 A	22-07-1997
		AU 678154 B	22-05-1997
		AU 3884493 A	08-11-1993
		CA 2133417 A	14-10-1993
		EP 0633940 A	18-01-1995
		FI 944550 A	30-11-1994
		JP 7505525 T	22-06-1995
		NO 943664 A	15-11-1994
		US 5792922 A	11-08-1998
WO 9623892 A	08-08-1996	US 5654495 A	05-08-1997
		CA 2212003 A	08-08-1996
		EP 0807182 A	19-11-1997
WO 9621022 A	11-07-1996	US 5614393 A	25-03-1997
		AU 4673596 A	24-07-1996
		CA 2207906 A	11-07-1996
		CN 1177379 A	25-03-1998
		EP 0801680 A	22-10-1997
		US 5789220 A	04-08-1998
WO 9311245 A	10-06-1993	AU 675923 B	27-02-1997
		AU 3228893 A	28-06-1993
		CA 2124673 A	10-06-1993
		EP 0616644 A	28-09-1994
		JP 7501701 T	23-02-1995
WO 9418337 A	18-08-1994	EP 0684998 A	06-12-1995
		JP 8506490 T	16-07-1996
WO 9411516 A	26-05-1994	AU 5407594 A	08-06-1994
		CA 2149223 A	26-05-1994
		EP 0668919 A	30-08-1995
		JP 8503364 T	16-04-1996
WO 9606936 A	07-03-1996	AU 3410295 A	22-03-1996
		BR 9509502 A	30-09-1997
		CA 2198222 A	07-03-1996
		EP 0778896 A	18-06-1997

INTERNATIONAL SEARCH REPORT

..formation on patent family members

Inte onal Application No
PCT/US 98/07179

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9606936 A		HU 76842 A	28-11-1997
		JP 10505237 T	26-05-1998
		PL 319103 A	21-07-1997

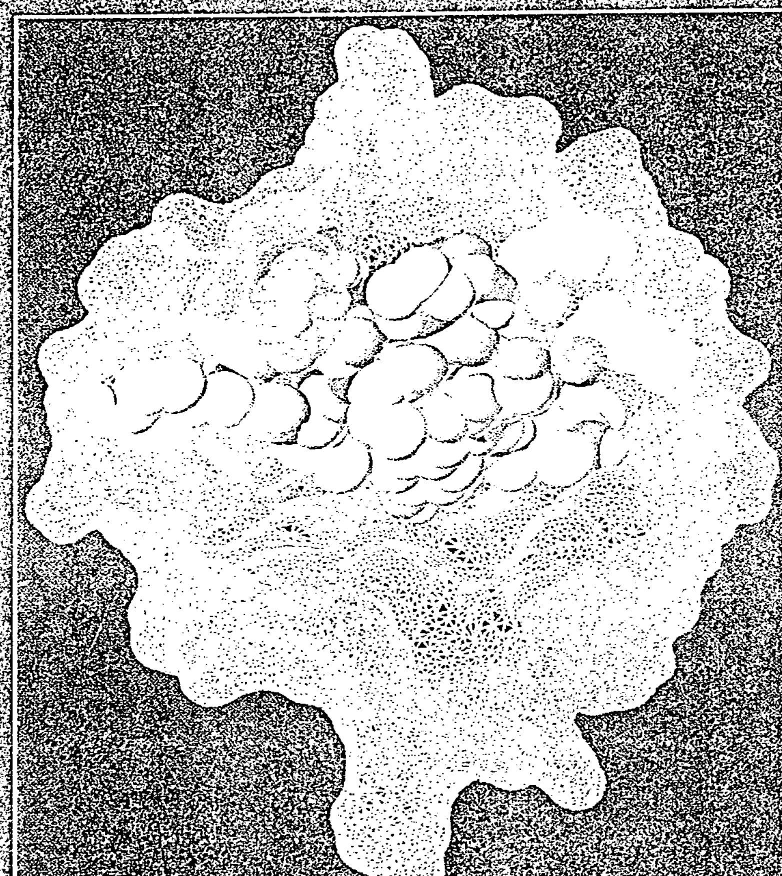
Form PCT/ISA/210 (patent family annex) (July 1992)

ISSN 0006-291X
This Number Completes Volume 255
Volume 255, Number 3, February 24, 1999

10

BBRC

Biochemical and Biophysical Research Communications



AP
ACADEMIC PRESS

Biochemical and Biophysical Research Communications 255, 575–579 (1999)
Article ID bbrc.1999.0235, available online at http://www.idealibrary.com on IDEAL®

Molecular Cloning and Functional Characterization of Rat Δ-6 Fatty Acid Desaturase

Tsunehiro Aki,[1] Yayoi Shimada, Katsuya Inagaki, Hirofumi Higashimoto, Seiji Kawamoto, Seiko Shigeta, Kazuhisa Ono, and Osamu Suzuki

Department of Molecular Biotechnology, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima 739-8527, Japan

Received January 6,1999

Mammalian cDNA fragments putatively encoding amino acid sequences characteristic of the fatty acid desaturase were obtained using expressed sequence tag (EST) sequence informations. These fragments were subsequently used to screen a rat liver cDNA library, yielding a 1573-bp clone. Expression of DNA fragment containing either of two possible open reading frames (nucleotide numbers 97-1431 and 148-1431) of the isolated clone in yeast led to the accumulation of γ-linolenic acid in the presence of exogenous linoleic acid. In this system, the addition of α-linolenic acid also resulted in the accumulation of its Δ-6 desaturated product whereas dihomo-γ-linolenic acid failed to be a substrate. These results indicate that the protein encoded by the rat cDNA is Δ-6 fatty acid desaturase, and the first 17 amino acids corresponding to the coding region 97-147 of the clone are not required to function in yeast. © 1999 Academic Press

Δ-6 Desaturase catalyzes the conversion of linoleic acid (LA, C18:2Δ-9, 12) to γ-linolenic acid (GLA, C18:3Δ-6, 9, 12) by inserting a double bond between carbon 6 and 7, in conjunction with cytochrome b_5-mediated electron transfer system in mammals. Since GLA and its elongation product, dihomo-γ-linolenic acid (DGLA, C20:3Δ-8, 11, 14), are barely detectable in mammalian cells, it is generally accepted that the Δ-6 desaturation step is rate-limiting (1). In this context, the activity of the Δ-6 desaturase is considered to affect directly to the cellular content of arachidonic acid (AA, C20:4Δ-5, 8, 11, 14) which is a Δ-5 desaturated product of DGLA. It is feasible to extend this aspects on the n-6 pathway to another pathway (n-3) where α-linolenic acid (ALA, C18:3Δ-9, 12, 15) is converted to eicosapentaenoic acid (EPA, C20:5Δ-5, 8, 11, 14, 17) through Δ-6 desaturation.

[1] To whom correspondence should be addressed. Fax: +81-824-22-7191. E-mail: aki@ipc.hiroshima-u.ac.jp.

AA is well-known as a precursor of a large family of eicosanoids that have multiple effects related to the regulation of e.g. blood pressure, inflammatory reactions, and platelet function (1–3). EPA exhibits antagonizing effect aganst AA metabolism, and *vise versa* (4, 5). Since the amounts and types of eicosanoids synthesized are partially determined by the availability of the fatty acid precursors, imbalance of these acids is suggested to contribute to numerous clinical symptoms. An early study indicated that affinity of the Δ-6 desaturase for ALA is greater than that for LA, implying that these fatty acids might not be metabolized in the same fashion (6). Therefore, the imbalance of the levels of fatty acid precursors could be due to the impaired activity of the Δ-6 desaturase on either of the two pathways. Indeed, depression of the Δ-6 desaturase activity, mainly reported on the n-6 pathway, is associated with various physiologic and pathophysiologic states including aging, diabetes, atopic dermatitis, cardiovascular disorders, and cancer (1, 7, 8). Also, the differences in nutritional and hormonal conditions influence the Δ-6 desaturase activity, resulting in the altered composition of intracellular fatty acids and membrane phospholipids (1, 9). Up to now, these observations have been led, in part, by tracing the activity of the enzyme, detected predominantly in the microsomal membrane fraction. However, molecular characterization of the membrane-bound desaturase protein especially in mammals has not been fully established.

Recently, genes coding for Δ-6 desaturases from the borage *Borago officinalis* (10) and the nematode *Caenorhabditis elegans* (11) and Δ-5 desaturases from *C. elegans* (12) and the fungus *Mortierella alpina* (13, 14) were identified. Mutual comparisons of their deduced amino acid sequences revealed the presence of highly conserved heme-binding motif and histidine boxes, located in same order, which appeared to be common in all desaturases of bacteria and plants (15). Taking advantage of the sequence informations on the desaturases in eukaryotes, we newly identified a rat liver

0006-291X/99 $30.00

cDNA encoding functional Δ-6 fatty acid desaturase, as reported here.

MATERIALS AND METHODS

General laboratory chemicals were purchased from Katayama Chemical (Osaka, Japan). Fatty acid standards were from Sigma Chemical Co. (St. Louis, MO). Reagents and enzymes for genetic manipulations were from Takara Shuzo (Kyoto, Japan), otherwise stated. Male BALB/c mice were obtained from Charles River Japan (Hiroshima, Japan).

Messenger RNAs were extracted from mouse liver by guanidinium thiocyanate method using QuickPrep mRNA Purification Kit (Amersham Pharmacia Biotech, Uppsala, Sweden). A cDNA pool was prepared from the mouse mRNAs by TimeSaver cDNA Synthesis Kit (Amersham Pharmacia Biotech) according to the manufacturer's instruction. Oligonucleotide primers of following sequences were synthesized for amplification of two apart regions of gene coding for desaturase-like protein by polymerase chain reaction (PCR); m3F, 5'-GTCAGGGTGCTGGAGAGCCACTGG-3'; m3R, 5'-GTAGTGTAGGCCGTGCTTCGCGC-3'; m5F, 5'-GATGCTACGGATGCCTTCCGTGC[illegible] m5R, 5'-TTCATGTCCTCAGCAGTCTTCTTC-3'. The cDNA fro[illegible]use liver was subjected to PCR reactions (LA PCR kit, Takara) with primer pairs, m3F and m3R, or m5F and m5R. Successfully amplified products, m3 and m5 respectively, were cloned on plasmid pGEM-T Easy Vector (Promega, Madison, WI), and the inserts were confirmed by DNA sequencing analyses.

Rat liver cDNA library constructed on λZAP II (#937507, Stratagene, La Jolla, CA) was probed with alkaline phosphatase-labeled m3 or m5 fragment, by which labeling of the probes, hybridization, and detection of hybrids were performed using AlkPhos Direct System (Amersham Pharmacia Biotech). In this protocol, hybridization and washing steps were done at 55°C. Plaques positively detected by either of the two probes were picked up, and the accuracy of the first screening was reevaluated with purified plaques by another round of plating and hybridization. The plasmids containing positive cDNA were recovered from selected λ clones by *in vivo* excision, and the insert was entirely sequenced on both strands of DNA.

Since one of the positive clones, r24, seemed to contain full-length cDNA of interest, a plasmid derived from clone r24 was used as a template for PCR amplification of regions which were deduced as open reading frames. Because two ATG sequences could be considered as putative translation initiation codons, two forward primers, r24aF, 5'-ACAAAGCTT**ATG**GGGAAGGGAGGTAACCAG-3' (corresponding to the first ATG indicated by boldface type) and r24bF, 5'-C[illegible]AAGCTT**ATG**CCCACCTTCCGCTGGGAG-3' (corresponding to t[illegible]econd ATG indicated by boldface type) were used to amplify the coding frames, r24a and r24b, respectively, and to generate *Hin* dIII site (underlined) adjucent to the ATG. A reverse primer r24R, 5'-TCTTCTAGA**TCA**TTTGTGGAGGTAGGCATC-3' (annealing to the complement of the stop codon indicated by boldface type) was used for each PCR reaction, generating *Xba* I site (underlined). The PCR products treated with *Hin* dIII and *Xba* I were inserted respectively to the yeast vector pYES2 (Invitrogen, San Diego, CA). It was confirmed by DNA sequencing analyses that the entire and flanking sequences of the inserts were as we designed. Transfer of the constructs into *Saccharomyces cerevisiae* strain INVSc1 (Invitrogen) was done by the lithium acetate method, and recombinant yeast cells were selected on uracil-deficient medium. The yeast cells were cultivated in a medium containing 4% raffinose, 0.7% yeast nitrogen base without amino acid, 1% tergitol type NP-40, 20 μg/ml histidine, 60 μg/ml leucine, and 40 μg/ml tryptophan at 28°C, overnight. The culture broth was supplemented with fatty acid substrate so as to be a final concentration of 0.5 mM, followed by further cultivation until cell density reached at 5×10^6 cells/ml. The expression of the transgene was performed by the addition of galactose to 2% (w/v) and an additional cultivation for 10 hr.

Culture broths were harvested, and total intracellular lipids were extracted with a mixture of chloroform/methanol (2:1, v/v). The lipid fraction was subjected to methyl-esterification with 10% hydrochloride in methanol. Fatty acid methyl esters were applied on a gas-liquid chromatograph (GC; model GC-17A, Shimadzu, Kyoto, Japan) equipped with a TC-70 capillary column (GL Science, Tokyo, Japan) and a flame ionization detector. The condition for GC analysis was as previously described (16). GC-mass spectrometry (GC-MS) analysis of the fatty acid methyl esters was performed using a MS-BU20 (JEOL, Tokyo, Japan) high-resolution mass spectrometer linked to a gas chromatograph (model MS-5890, Hewlett Packard) equipped with the TC-70 column as the sample inlet, and operated in the electron impact mode at 70 eV. Comparison of the mass spectra of authentic standards and interest peaks in total ion chromatogram was done by visual- and computer-based examinations.

RESULTS

A nucleotide sequence corresponding to the highly conserved region (indicated by dotted line in Fig. 1) in Δ-6 desaturase from *C. elegans* (11) was used as a query to search databases for related sequences in mammals. When the database of mouse expressed sequence tag (EST) at DNA Data Bank of Japan was searched using both the BLAST and FASTA algorithms, several entries registering DNA sequences partially homologous to the query were retrieved. The nucleotide sequence in 5'-region of one (GenBank accession number W53753) of the ESTs was then used to further search the database and found to be partially overlapped with another EST clone (AA512429). By similar sequential searchs toward 5'-end of a putative desaturase gene in mouse, a clone (AA036321) overlapped with the AA512429 was found, and the AA036321 led us to an additional clone (AA250162). Nucleotide sequence of the AA250162 could be translated to the amino acid sequence bearing partial resemblance to that of N-terminal domain of previously characterized desaturases.

Based on the sequence informations from ESTs W53753 and AA250162, we made two non-overlapping DNA fragments, m3 (3'-region) and m5 (5'-region), respectively, by PCR with a mouse liver cDNA pool as a template. These fragments were used as hybridization probes for isolation of entire coding region of desaturase gene from rat liver cDNA library. We elected rat, instead of mouse, as a source of the target gene, because the desaturases had been best-characterized biochemically in rat, which included a report of partial purification of linoleoyl-CoA desaturase (17; see Discussion). The condition for hybridization was set at medium stringency making allowance for differences in animal species. As a result of screening, five individual clones were isolated as positives to the probe m3 and only one of them, termed r24, was hybridized also with m5 probe. Sequencing of these clones revealed that the clone r24 had a cDNA insert of 1573 basepairs (bp) in length (GenBank accession number AB021980), and

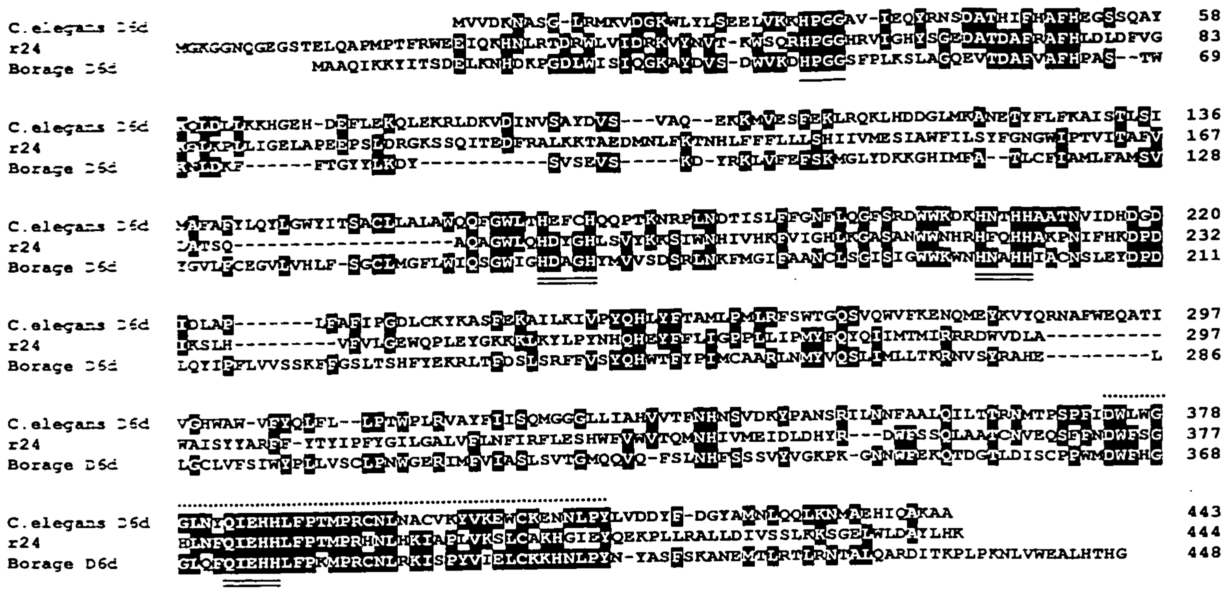


FIG. 1. Composite alignments of the amino acid sequence deduced from the rat cDNA clone r24 with Δ-6 desaturases from other sources. Borage D6d, *B. officinalis* Δ-6 desaturase (GenBank accession number U79010); C.elegans D6d, *C. elegans* Δ-6 desaturase (AF031477). Nucleotide sequence corresponding to the highly conserved region (indicated by dotted line above the sequences) in *C. elegans* Δ-6 desaturase was used as a query to search databases, as mentioned in the text. Identical residues are boxed, and the conserved heme-binding motif and three histidine boxes are marked with single and double underlines, respectively.

cDNAs in remaining four clones (all of them are less than 500 bp in length) were corresponded to 3′-region of r24 (data not shown), supporting the failure of hybridizing with probe m5. A putative protein encoded by the clone r24 seemed to be a rat homolog of the protein from the mouse EST AA250162 (96.5% identity in 491 nucleotides overlap). Thus, we chose the clone r24 for further characterization.

Two ATG initiation codons (nucleotide numbers 97-99 and 148-150) were found in the sequence of 5′-terminal domain of the r24 cDNA and those were placed in-frame. According to the Kozak consensus sequence, AGXXATGG, that has been advocated as favored sequence for eukaryotic initiation sites (18), the first of the two initiation codons is credible. If this is the case, the cDNA contains a coding frame of 1335-bp long including a TGA termination codon (nucleotide numbers 1429-1431), which can be translated into 444 amino acid polypeptide. Comparisons of the deduced amino acid sequence of r24 with Δ-6 desaturases from *C. elegans* and *B. officinalis* showed homology scores of 27.9% and 26.4%, respectively (Fig. 1). It is noted in Fig. 1 that a typical heme-binding motif, HPGG (19), and three histidine boxes highly conserved within fatty acid desaturases (15) are present in the r24 sequence as well as the others. At the third histidine box, the first histidine residue in the conventional motif, HXXHH, was substituted with glutamine. This variance had occured in Δ-6 and Δ-5 desaturases from fungus, plant, and lower animal (10–14).

For functional analysis of the clone r24, two possible coding regions, named r24a and r24b (nucleotide numbers 97-1431 and 148-1431, respectively) were amplified by PCR, and respective expression plasmids were constructed on the yeast vector pYES2. The PCR products were located at just downstream of the galactose-inducible *GAL1* promoter on each construct. After obtaining yeast transformants carrying pYES2/r24a, pYES2/r24b, or pYES2 (control), cells were cultivated, supplemented by the addition of substrate LA (C18:2Δ-9, 12), and induced in the presence of galactose. Aliquots of cells in the induced culture broth were taken for analyses of the intracellular fatty acid composition by GC, and the resultant chromatograms of fatty acid methyl esters were shown as Fig. 2. A novel peak, which was not apparent in the case of control (Fig. 2A), was detected in charts from both induced pYES2/r24a (peak 6 in Fig. 2B) and pYES2/r24b (data not shown). Similarly, when the substrate LA was replaced with ALA (C18:3Δ-9, 12, 15), a peak additional to the background level in the control case (Fig. 2C) was found in a GC profile obtained from the yeast transformed with either pYES2/r24a (peak 8 in Fig. 2D) or pYES2/r24b (data not shown). We confirmed that these and other additional peaks did not appear when the yeast carrying pYES2/r24a or pYES2/r24b was not induced by galactose or was supplemented with none of exogenous fatty acids. Comparisons of the retention times of the newly yielded peaks with those of authentic standards have anticipated that the fatty

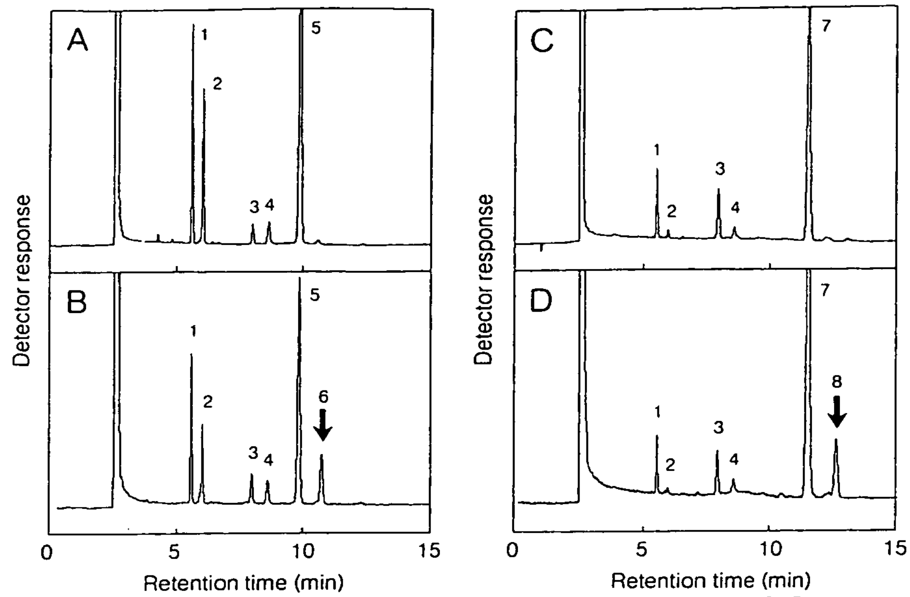


FIG. 2. GC analyses of methyl esterified fatty acids from the induced yeast cells containing pYES2 (A and C) or pYES2/r24a (B and D). Before the induction with galactose, LA (peak 5 in A and B) or ALA (peak 7 in C and D) was added to be incorporated to the cells. The peaks that found only in the case of pYES2/r24a were indicated by arrows (peak 6 in B and peak 8 in D). Identities of other peaks were determined by comparing their retention times with those of authentic standards. Peaks 1, C16:0; 2, C16:1Δ-9; 3, C18:0; 4, C18:1Δ-9.

acids giving the peaks 6 and 8 are GLA (C18:3Δ-6, 9, 12) and *cis*-3, 6, 9, 12-octadecatetraenoic acid (C18: 4Δ-6, 9, 12, 15), which are the Δ-6 desaturation products of LA and ALA, respectively. These prospects were positively supported by definitive assignments of the compounds in peaks 6 and 8 by GC-MS analyses (data not shown). In a separate experiment, when DGLA (C20:3Δ-8, 11, 14), a substrate of Δ-5 desaturase, was added to our expression system, no extra peak was observed in chromatograms from the r24a/r24b recombinants, compared to the negative control. Taken together, the recombinant yeast containing the inducible r24 cDNA had gained function of Δ-6, but not Δ-5, fatty acid desaturation.

DISCUSSION

Here we isolated a rat liver cDNA coding for the Δ-6 fatty acid desaturase. Although the cDNA, r24, was successfully expressed in yeast, we could not predict the actual ATG initiation codon corresponding to a methionine residue at the amino terminus of the native desaturase protein. This is because, in our study, no significant differences have not been detected between the two lines of expression analyses on r24a and r24b. This observation suggested no other than the needless of the first 17 amino acids in the protein expressed from r24a to function in yeast although this portion might be indispensable in rat. Another set of experiments including the purification of the native Δ-6 desaturase is essential to clarify this point and is being undertaken.

Okayasu, *et al.* (17) described a purification of rat liver linoleoyl-CoA desaturase that was capable of converting linoleoyl-CoA to γ-linolenoyl-CoA *in vitro*. The apparent molecular weight of this enzyme (66 kD) obviously differs from either molecular weights calculated from the deduced amino acid sequence of r24a (52.4 kD) or r24b (50.7 kD). This inevitably raises a possibility of the presence of more than two types of the enzymes taking charge of the Δ-6 fatty acid desaturation. Sprecher and his colleagues have proposed a novel pathway, docosapentaenoic acid to docosahexaenoic acid via Δ-6 desaturation, for the biosynthesis of polyunsaturated fatty acids (20, 21). However, the putative involvement of a single cycle of peroxisomal β-oxidation in this pathway is under a critical reevaluation, excluding also the necessity of the proposed Δ-6 desaturation step (22). A metabolic study by Christiansen, *et al.* (23) suggested that liver microsomes might contain separate enzymes for desaturation of LA and ALA. Their observations, however, seem to be inconsistent with a result of competitive study using fatty acid tracers (24), implying that a single enzyme may govern desaturating fatty acids at Δ-6 position. To date, no clear conclusions have been made whether multiple forms of Δ-6 desaturase exist.

In relation to these pending questions, we are attempting to isolate and characterize a full-length cDNA corresponding to the EST clones W53753, AA512429, and AA036321 since nucleotide sequences of these clones can be translated into amino acid sequences that are significantly homologous, but not

identical, to the sequence from our clone r24 (data not shown). This gene may encode an isoform of the Δ-6 desaturase, which is dominantly expressed in tissues other than liver or at the different developmental stages. This assumption does not contradict the facts that we were unable to isolate a liver cDNA whose sequence is matched with the probe m3 (from W53753), and these ESTs are derived from embryo and mammary gland. Otherwise, a protein encoded by this gene may be one of other desaturases, for example, Δ-5 desaturase which has not yet been identified in mammals. In either case, the cloning of the mammalian desaturase gene(s) will accelerate to elucidate the molecular mechanisms on the regulation of various cellular events by the enzyme possibly through the alteration of physical state of membrane lipids and of the level of pooled precursors for signal transducers.

ACKNOWLEDGMENTS

We thank Dr. Dai Hirata for his helpful advice on management of yeast. This work was supported in part by a Grant-in-Aid for Encouragement of Young Scientists (#10760052) from the Ministry of Education, Science and Culture, Japan.

REFERENCES

1. Cook, H. W. (1996) *in* Biochemistry of Lipids, Lipoproteins and Membranes (Vance, D. E., and Vance, J., Eds.), pp. 129–152 Elsevier Science, Amsterdam.
2. Marx, J. L. (1982) *Science* **215,** 1380–1383.
3. Samuelsson, B., Goldyne, E., Granström, E., Hamberg, M., Hammarström, S., and Malmsten, C. (1978) *Annu. Rev. Biochem.* **47,** 997–1029.
4. Li, B-. Y., Birdwell, C., and Whelan, J. (1994) *J. Lipid Res.* **35,** 1869–1877.
5. Whelan, J. (1996) *J. Nutr.* **126,** 1086S–1091S.
6. Brenner, R. R. and Peluffo, R. O. (1966) *J. Biol. Chem.* **241,** 5213–5219.
7. German, J. B., Dillard, C. J., and Whelan, J. (1995) *J. Nutr.* **126,** 1076S–1080S.
8. Horrobin, D. F. (1992) *Prog. Lipid Res.* **31,** 163–194.
9. Brenner, R. R. (1981) *Prog. Lipid Res.* **20,** 41–47.
10. Sayanova, O., Smith, M. A., Lapinskas, P., Stobart, A. K., Dobson, G., Christie, W. W., Shewry, P. R., and Napier, J. A. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 4211–4216.
11. Napier, J. A., Hey, S. J., Lacey, D. J., and Shewry, P. R. (1998) *Biochem. J.* **330,** 611–614.
12. Michaelson, L. V., Napier, J. A., Lewis, M., Griffiths, G., Lazarus, C. M., and Stobart, A. K. (1998) *FEBS Lett.* **439,** 215–218.
13. Michaelson, L. V., Lazarus, C. M., Griffiths, G., Napier, J. A., and Stobart, A. K. (1998) *J. Biol. Chem.* **273,** 19055–19059.
14. Knutzon, D. S., Thurmond, J. M., Huang, Y. -S., Chaudhary, S., Bobik, E. G., Chan, G. M., Kirchner, S. J., and Mukerji, P. (1998) *J. Biol. Chem.* **273,** 29360–29366.
15. Shanklin, J., Whittle, E., and Fox, B. G. (1994) *Biochemistry* **33,** 12787–12794.
16. Aki, T., Matsumoto, Y., Morinaga, T., Kawamoto, S., Shigeta, S., Ono, K., and Suzuki, O. (1998) *J. Ferment. Bioeng.* **86,** 504–507.
17. Okayasu, T., Nagao, M., Ishibashi, T., and Imai, Y. (1981) *Arch. Biochem. Biophys.* **206,** 21–28.
18. Kozak, M. (1981) *Nucleic Acids Res.* **9,** 5233–5262.
19. Napier, J. A., Sayanova, O., Stobart, A. K., and Shewry, P. R. (1997) *Biochem. J.* **328,** 717–720.
20. Voss, A., Reinhart, M., Sankarappa, S., and Sprecher, H. (1991) *J. Biol. Chem.* **266,** 19995–20000.
21. Sprecher, H., Luthria, D. L., Mohammed, B. S., and Baykousheva, S. P. (1995) *J. Lipid Res.* **36,** 2471–2477.
22. Infante, J. P. and Huszagh, V. A. (1998) *FEBS Lett.* **431,** 1–6.
23. Christiansen, E. N., Lund, J. S., Rørtveit, T., and Rustan, A. (1991) *Biochem. Biophys. Acta* **1082,** 57–62.
24. Geiger, M., Mohammed, B. S., Sankarappa, S., and Sprecher, H. (1993) *Biochem. Biophys. Acta* **1170,** 137–142.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 274, No. 1, Issue of January 1, pp. 471–477, 1999
Printed in U.S.A.

Cloning, Expression, and Nutritional Regulation of the Mammalian Δ-6 Desaturase*

(Received for publication, September 21, 1998)

Hyekyung P. Cho, Manabu T. Nakamura, and Steven D. Clarke‡

From the Program of Nutritional Sciences and the Institute for Cellular and Molecular Biology, The University of Texas-Austin, Austin, Texas 78712

Arachidonic acid (20:4(n-6)) and docosahexaenoic acid (22:6(n-3)) have a variety of physiological functions that include being the major component of membrane phospholipid in brain and retina, substrates for eicosanoid production, and regulators of nuclear transcription factors. The rate-limiting step in the production of 20:4(n-6) and 22:6(n-3) is the desaturation of 18:2(n-6) and 18:3(n-3) by Δ-6 desaturase. In this report, we describe the cloning, characterization, and expression of a mammalian Δ-6 desaturase. The open reading frames for mouse and human Δ-6 desaturase each encode a 444-amino acid peptide, and the two peptides share an 87% amino acid homology. The amino acid sequence predicts that the peptide contains two membrane-spanning domains as well as a cytochrome b_5-like domain that is characteristic of nonmammalian Δ-6 desaturases. Expression of the open reading frame in rat hepatocytes and Chinese hamster ovary cells instilled in these cells the ability to convert 18:2(n-6) and 18:3(n-3) to their respective products, 18:3(n-6) and 18:4(n-3). When mice were fed a diet containing 10% fat, hepatic enzymatic activity and mRNA abundance for hepatic Δ-6 desaturase in mice fed corn oil were 70 and 50% lower than in mice fed triolein. Finally, Northern analysis revealed that the brain contained an amount of Δ-6 desaturase mRNA that was several times greater than that found in other tissues including the liver, lung, heart, and skeletal muscle. The RNA abundance data indicate that prior conclusions regarding the low level of Δ-6 desaturase expression in nonhepatic tissues may need to be reevaluated.

Long chain polyunsaturated fatty acids such as 20:4(*n*-6) and 22:6(*n*-3) play pivotal roles in a number of biological functions including brain development, cognition, inflammatory responses, and hemostasis (1–4). Over 30% of the fatty acid in brain phospholipid consists of 20:4(*n*-6) and 22:6(*n*-3), and approximately 50% of the fatty acid in the retina is 22:6(*n*-3) (5, 6). An inadequate availability of 20:4(*n*-6) is associated with impaired nerve transmission, reduced eicosanoid synthesis, and impaired fetal growth (7–9). Recently, premature infants were found to have reduced cognitive development, apparently because they could not synthesize adequate quantities of 22:6(*n*-3) to meet the biological demands for proper retina function (1, 10). In addition to being vital components of membrane phospholipids and functioning in key steps of cell signaling, 20- and 22-carbon polyunsaturated fatty acids govern the expression of a wide array of genes, including those encoding proteins involved with lipid metabolism, thermogenesis, and cell differentiation (11–14).

The availability of 20- and 22-carbon (*n*-6) and (*n*-3) polyenoic fatty acids is greatly dependent upon the rate of desaturation of 18:2(*n*-6) and 18:3(*n*-3) by Δ-6 desaturase (15). Δ-6 Desaturase is a microsomal enzyme (15) and is thought to be a component of a three-enzyme system that includes NADH-cytochrome b_5 reductase, cytochrome b_5, and Δ-6 desaturase (16). The enzymatic activity for Δ-6 desaturase is reportedly low in most tissues except the liver (16). Consequently, the liver has been considered the primary site for the production of long chain polyenoic fatty acids (17, 18). Numerous dietary studies indicate that hepatic Δ-6 desaturase activity is induced by diets low in essential fatty acids and suppressed by diets rich in vegetable or marine oils (19, 20). In addition, Δ-6 desaturase activity is induced by peroxisome proliferators and by the administration of insulin to diabetic rats (21, 22). Unfortunately, defining the molecular determinants of Δ-6 desaturase activity, as well as characterizing its developmental pattern and tissue distribution, has been hampered by the fact that Δ-6 desaturase has been neither cloned nor reproducibly purified. Therefore, our objective was to clone the mammalian Δ-6 desaturase and utilize the cDNA to examine the tissue distribution and nutritional regulation of Δ-6 desaturase mRNA.

EXPERIMENTAL PROCEDURES

Cloning of the Mouse Δ-6 Desaturase cDNA—A murine cDNA (GenBank accession number W53753) displaying high homology to the amino acid sequence of Δ-6 desaturase from *Synechocystis* sp. was acquired and sequenced. Subsequently, a 23-base oligonucleotide primer (5′-CTTGGCATCGTGGGGAAGAGGTG-3′) specific for the 5′ end of murine cDNA W53753 was synthesized and utilized to screen a mouse adaptor-ligated liver cDNA library (Marathon-Ready cDNA; CLONTECH) by rapid amplification of cDNA ends-PCR.[1] The PCR conditions consisted of an initial denaturation step of 94 °C for 1 min, followed by 5 cycles of 94 °C for 10 s and 72 °C for 4 min, 5 cycles of 94 °C for 10 s and 70 °C for 4 min, and, finally, 20 cycles of 94 °C for 10 s and 68 °C for 4 min. The resulting rapid amplification of cDNA ends-PCR product was cloned into pBluescript (Stratagene) and sequenced by the dideoxy chain termination method (23).

The nucleotide sequence of the PCR product was utilized to BLAST search the mouse EST database. Two mouse cDNAs (GenBank accession numbers AA237892 and AA250162) possessing 100% nucleotide homology with our PCR product were identified and acquired from Genome Systems. Clone AA250162 contained two possible AUG start codons, and the EST cDNA AA237892 contained an apparent stop codon. The two EST cDNAs were fused at the *Sty*I restriction site, and

* This work was supported by the National Institutes of Health Grant DK 52573 and the sponsors of the M. M. Love Chair in Nutritional, Cellular and Molecular Sciences at the University of Texas-Austin (S. D. C.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: 115 GEA, The University of Texas at Austin, Austin, TX 78712. Tel.: 512-232-1537; Fax: 512-471-5630; E-mail: stevedclarke@mail.utexas.edu.

[1] The abbreviations used are: PCR, polymerase chain reaction; ORF, open reading frame; EST, expressed sequence tag; CHO, Chinese hamster ovary; kb, kilobase pair(s).

the product was inserted into the cytomegalovirus promoter expression vector pcDNA3.1 (Invitrogen). Sequence analysis and prediction of amino acid sequence were performed using MacDNASIS pro (Hitachi), and a translation initiator codon was determined based on Kozak's rule (24).

Cloning of the Human Δ-6 Desaturase cDNA—Using the nucleotide sequence of mouse liver Δ-6 desaturase, the EST human database was searched for a human homologue cDNA. The search identified a highly homologous human brain EST cDNA (GenBank accession number Z44979), which was purchased from Genome Systems and sequenced for verification. The 5′ end of the human cDNA was extended by PCR using a human brain cDNA library (Marathon-Ready cDNA; CLONTECH). The forward oligonucleotide primer (5′-AGACTGGCAGCATGGGGAAG-3′) was prepared using the 5′ end of the mouse Δ-6 desaturase and designed to include the putative start codon. The reverse primer (5′-CATGGTGGGGAAGAGGTGGTG-3′) was prepared from the sequence derived from human Z44979 cDNA.

Expression of the Mouse Δ-6 Desaturase—Cellular expression of the mouse Δ-6 desaturase was performed in rat primary hepatocytes and CHO cells. Rat primary hepatocytes were isolated by collagenase perfusion and allowed to attach to a collagen-coated 60-mm culture plate in 3 ml of Waymouth 752 medium supplemented with 0.5% fetal bovine serum and 1 μM insulin and dexamethasone (25). After a 6-h attachment period, the cells were washed with phosphate-buffered saline and transfected with 6 μg of the mouse Δ-6 desaturase expression plasmid or the pcDNA3.1 expression vector alone using 6.6 μl of Lipofectin per μg of DNA (Life Technologies, Inc.). Transfection was conducted by adding the mixture of Lipofectin and DNA in the absence of fetal bovine serum. After the 12-h transfection period, the medium was replaced with the one containing either 200 μM albumin-bound 18:2, n-6 (molar ratio of fatty acid to albumin, 4:1) or albumin alone. CHO cells were grown in Kaighn's modification of Ham's F-12 medium supplemented with 10% fetal bovine serum in a 25-cm^2 flask. At 80% confluence, the serum-containing medium was removed, and cells were washed with phosphate-buffered saline for transfection. A mixture of 2 μg of the mouse Δ-6 desaturase expression plasmid, 12 μl of LipofectAMINE, and 8 μl of Plus reagent (Life Technologies, Inc.) was added to cells without serum for 4 h. Subsequently, 10% serum was added to the transfection media for 8 h. After a total 12-h transfection period, the CHO cells were treated with either 200 μM albumin-bound 18:3(n-3), 20:3(n-6), or albumin alone. The hepatocytes and CHO cells were incubated with the treatment medium for 24 h and then used for fatty acid analysis.

Fatty Acid Extraction and Analysis—Cellular fatty acid was extracted by saponifying fatty acids using 1 ml of 30% KOH and 1 ml of ethanol. Fatty acids from the treatment medium were also extracted and analyzed after 24 h of incubation. Heptadecanoic acid was added to the saponification mixture as an internal standard. After saponification, the nonsaponifiable lipids were removed by extraction with petroleum ether. Subsequently, the solution was acidified, and the fatty acids were extracted with petroleum ether. The extract was dried under nitrogen, and the residue was methylated using 14% boron trifluoride in methanol (Sigma). Methylated fatty acids were separated and quantified by gas chromatography using a fused silica glass capillary column (50 m × 530 μm internal diameter; Quadrex). The column temperature program was composed of an initial hold at 140 °C for 5 min, ramping at 5 °C per min to 220 °C, and a final hold at 220 °C for 7 min. The injector temperature was 250 °C, and the flame ionization detector temperature was 260 °C.

Nutritional Regulation of Δ-6 Desaturase Expression—Male BALB/c mice were fed a high-glucose, fat-free diet for a 7-day adaptation period. After this period, the fat-free diet was supplemented with either 10% corn oil or 10% triolein (Sigma; 99% purity), and the mice (n = 4 mice/group) were fed for an additional 5 days. Liver tissues were removed, and microsomes were isolated by differential centrifugation. One g of liver was homogenized in 4 ml of homogenization buffer containing 50 mM potassium phosphate, pH 7.4, and 0.25 M sucrose. After a 10-min centrifugation of the homogenate at 10,000 × g, the resulting supernatant was spun at 100,000 × g for 60 min to isolate a microsomal pellet. After resuspending the pellet in homogenization buffer, 3 mg of microsomal protein were incubated in a 37 °C shaking water bath for 5 min with 1 ml of reaction mixture including 1.2 mM NADH, 3.6 mM ATP, 0.5 mM coenzyme A, 4.8 mM $MgCl_2$, 72 mM phosphate buffer, pH 7.4, and 50 nmol of 1-^{14}C-labeled 18:2(n-6). The reaction was stopped by adding saponification reagent, and fatty acids were saponified and methylated as described above. Radioactive 18:2(n-6) and 18:3(n-6) were separated by silver nitrate-impregnated thin layer chromatography. The radioactivity was quantified using an Ambis radio-imager. Δ-6 Desaturase enzyme activity is expressed as the percentage of 18:2(n-6) converted to 18:3(n-6) per mg of protein/min.

RNA Extraction and Northern Blot Analysis—Total RNA was isolated from the liver of mice in the dietary study using the phenol-guanidinium isothiocyanate method (26). Twenty μg of total RNA were size-fractionated on a 1% formaldehyde gel and then transferred to a Zeta probe nylon membrane (Bio-Rad). The mouse Δ-6 desaturase probe was prepared by incorporating [^{32}P]dCTP by PCR. The forward primer was 5′-GGACATAAAGAGCCTGCATG-3′, and the reverse primer was 5′-ACTGGAAGTACATAGGGATG-3′. The Northern membrane of human tissues was purchased from Invitrogen. The radiolabeled probe for the human tissue blot was a 200-base pair PCR fragment of human Δ-6 desaturase using primers of 5′-GGCAAGAACTCAAAGATCAC-3′ and 5′-GAGAGGTAGCAAGAACAAAG-3′. The autoradiographic signal was quantified using Instant Imager (Packard).

RESULTS

Cloning Mouse and Human Δ-6 Desaturase—The mouse EST database was searched for mammalian homologues using the amino acid sequence for Δ-6 desaturase from the photosynthetic cyanobacterium *Synechocystis* sp. (27). A mouse cDNA that had a 60% similarity to a 46-amino acid sequence of *Synechocystis* Δ-6 desaturase was identified. A 1508-base pair cDNA sequence for mouse liver Δ-6 desaturase was acquired using a combination of ligation-mediated PCR screening of a mouse liver cDNA library and BLAST searches of the mouse EST database (Fig. 1*A*). Sequence analysis revealed the presence of two in-frame methionine codons located at positions 75 and 126. In addition, a TGA termination sequence was identified at position 1407. Kozak's rule, which predicts that the favored eukaryotic translation initiation sequence resides in the sequence of AXXATGG (24), indicated that the first of the two ATG codons was the preferred initiation codon for the putative Δ-6 desaturase. The apparent ORF between the first ATG codon and the TGA termination codon predicted a peptide consisting of 444 amino acids and having a size of 52.2 kDa. The human brain cDNA homologue for Δ-6 desaturase contains an initiation codon and a termination codon that are perfectly aligned with the initiation and termination codons of the mouse cDNA (Fig. 1*A*). Moreover, the amino acid sequence derived from the ORF revealed that 87% of the amino acid sequence for the mouse and human homologues was identical, and 96% of the sequence had similarity (Fig. 1*B*). A search of the Swiss Protein Database indicated that the putative Δ-6 desaturase sequence was unique and shared very little amino acid homology with any other mammalian proteins including the murine stearoyl-CoA desaturase (Δ-9 desaturase) (28).

Structural Characteristics of Mammalian Δ-6 Desaturase—The enzymatic activity of mammalian Δ-6 desaturase is associated with the microsomal membrane fraction (15). Consistent with such membrane involvement, the predicted amino acid sequence for Δ-6 desaturase indicated that the peptide contains 52% nonpolar amino acids, and a hydropathy profile revealed the presence of two membrane-spanning domains that are characteristic of membrane-anchored proteins (Figs. 1*B* and 2). In addition, the amino terminus of the Δ-6 desaturase peptide contains a hydrophilic domain of 54 amino acids that is highly homologous with the heme-binding domain of cytochrome b_5 (Fig. 3*A*). This cytochrome b_5-like domain is also found in the Δ-6 desaturases from *Borago officinalis* (29) and *Caenorhabditis elegans* (30) (Fig. 3*A*). The His^{53} and His^{75} residues located within this domain of the mammalian Δ-6 desaturase are exactly aligned with the two heme-binding histidines in cytochrome b_5 (31). Moreover, these two histidines are surrounded by charged amino acids that may contribute to the stabilization of the heme-histidine complex (31). In addition, the sequence ^{53}HPGG^{56} predicts the existence of a dramatic β-turn that may render His^{53} more accessible to heme iron binding (31).

A second noteworthy feature of the mammalian Δ-6 desaturase is the presence of three histidine-rich regions (Fig. 3*B*

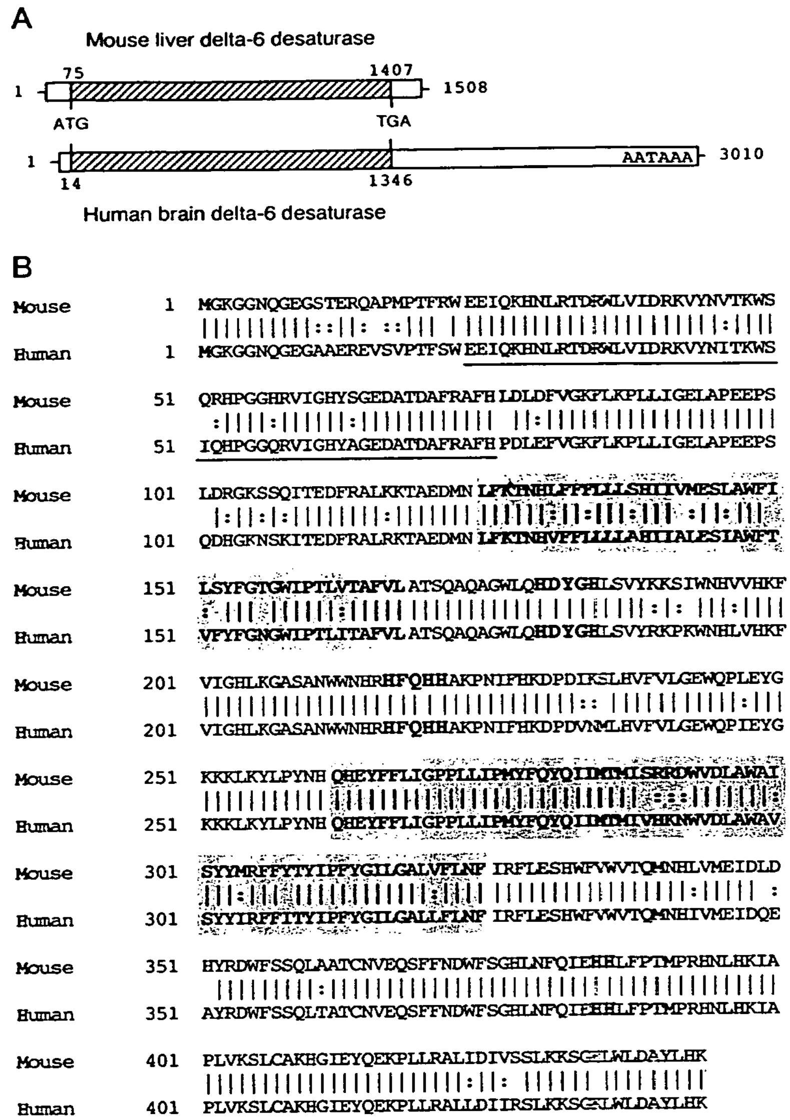


FIG. 1. **Alignment of the predicted amino acid sequences for mouse and human Δ-6 desaturase.** *A*, schematic diagram of the ORF and untranslated regions for mouse and human Δ-6 desaturase. The *hatched box* indicates an ORF of 1332 nucleotides, and the *open boxes* represent untranslated regions. The human cDNA contains a polyadenylation signal AATAAA at the 3′ end. *B*, a comparison of the amino acid sequences for mouse and human Δ-6 desaturase predicted by the nucleotide sequence of the ORFs. Both mouse and human ORFs encode 444 amino acids. Identical amino acids are paired by *vertical lines*, and conserved amino acids are matched by *colons*. The cytochrome b_5-like domain is *underlined*. Transmembrane domains are shown in *shaded areas*, and three histidine-rich domains are in *bold*.

Regions I (HX_3H) and II (HX_2HH) are located between the two transmembrane domains, and region III (HH) is located near the carboxyl terminus of the peptide. These histidine-rich regions are also found in plant membrane desaturases and mammalian stearoyl-CoA desaturase and reportedly bind non-heme iron that is required for enzymatic activity (32).

Expression of Δ-6 Desaturase—The predicted structural characteristics of the mouse and human peptides strongly suggested that the cDNAs did in fact correspond to mammalian Δ-6 desaturase. To confirm this conclusion, the ORF for the mouse Δ-6 desaturase was expressed in primary cultures of rat hepatocytes and in CHO cells. Fatty acid analysis revealed that hepatocytes transfected with the vector containing the Δ-6 desaturase ORF were capable of synthesizing the Δ-6 desaturase product 18:3(n-6) from 18:2(n-6) (Fig. 4*A*). On the other hand, hepatocytes transfected with vector alone produced no detectable 18:3(n-6) product (Fig. 4*B*). Similarly, CHO cells expressing Δ-6 desaturase readily converted the Δ-6 desaturase substrate 18:3(n-3) to the Δ-6 desaturase product 18:4(n-3), whereas nontransfected CHO cells were unable to produce detectable levels of 18:4(n-3) (Fig. 4, *C* and *D*). In contrast, providing CHO cells with the Δ-5 desaturase substrate 20:3(n-6) did not lead to the production of the Δ-5 desaturase product 20:4(n-6) (data not shown). These data conclusively demonstrate that the mouse and human ORFs do in fact encode mammalian Δ-6 desaturase.

Nutritional Regulation of Δ-6 Desaturase Expression—The enzymatic activity of Δ-6 desaturase increases when animals are fed an essential fatty acid-deficient diet, whereas it decreases when polyunsaturated fatty acids are ingested (16, 19, 20). Using the mouse cDNA for Δ-6 desaturase, we have found that the suppression of hepatic Δ-6 desaturase enzymatic activity associated with the ingestion of polyunsaturated fat (*i.e.* corn oil) is paralleled by a comparable reduction in Δ-6 desaturase mRNA abundance (Fig. 5, *A* and *B*). Interestingly, whereas the dominant transcript of hepatic Δ-6 desaturase is approximately 4.0 kb in size, the mouse liver also contains a minor transcript that is approximately 2.2 kb (Fig. 5*B*). Both transcripts appeared to be suppressed by dietary corn oil to the same degree. In addition, hybridizing the Northern blot with sequences from the 5′, middle, and 3′ regions of the Δ-6 desaturase ORF yielded the same outcomes with respect to the abundance and dietary response of the 2.2-kb transcript (data not shown). The reason for these two different transcripts remains unknown.

Δ-6 Desaturase mRNA Distribution in Human Tissues—

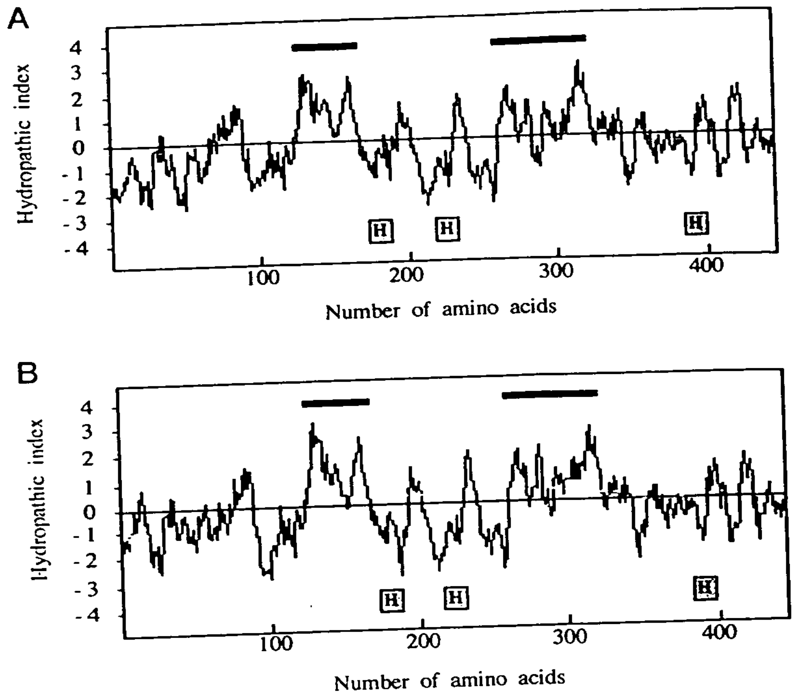


FIG. 2. **Hydropathy profile of mouse (*A*) and human (*B*) Δ-6 desaturase.** The hydropathic pattern for Δ-6 desaturase was plotted using the method of Kyte-Doolittle, and the amino acid sequences were predicted by the respective ORFs. *Bars*, the transmembrane regions. *Boxed H*, locations of histidine-rich regions.

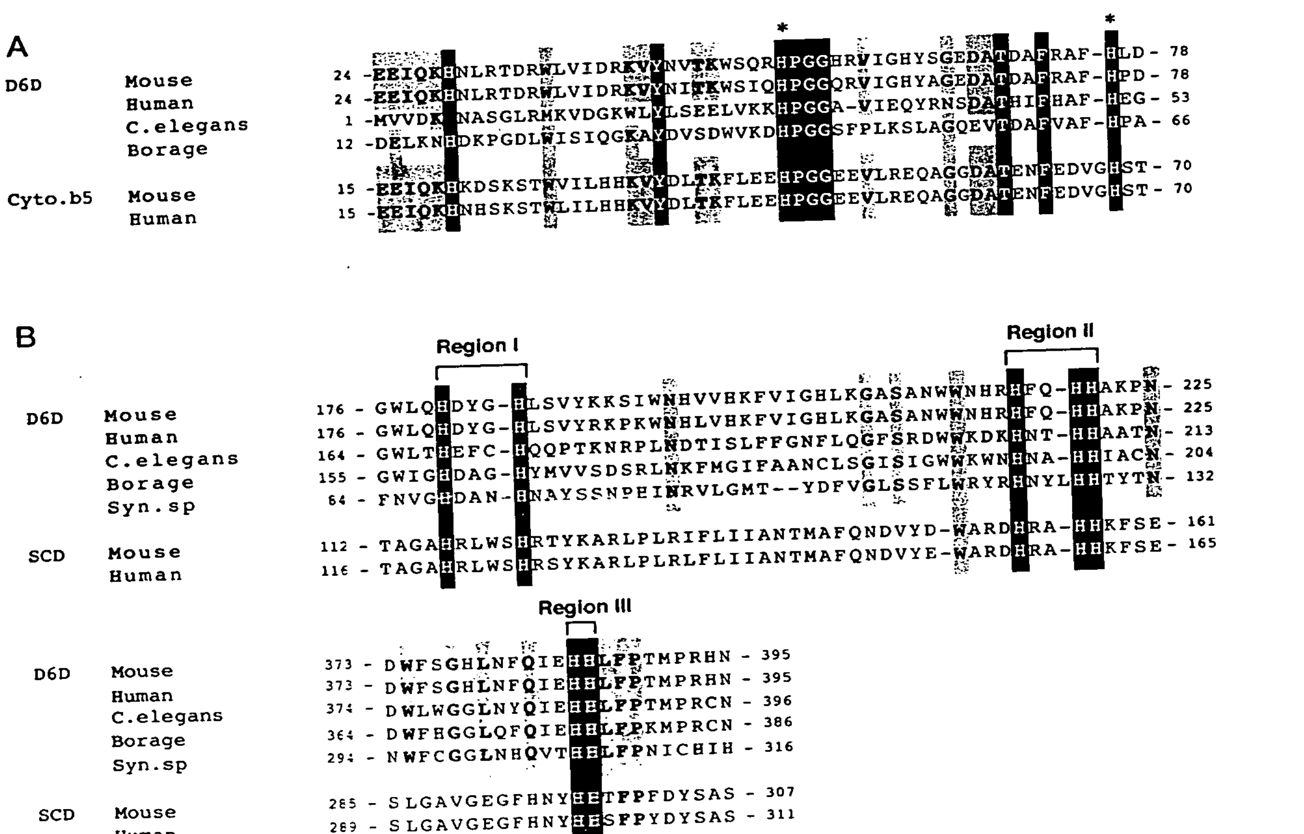


FIG. 3. **A comparison of the cytochrome b_5-like and histidine-rich domains for mammalian and nonmammalian Δ-6 desaturases.** *A*, a comparison of the cytochrome b_5-like domain for mammalian and nonmammalian Δ-6 desaturase (*D6D*). A comparison of the amino acid sequence within the cytochrome b_5-like domain of mouse, human, plant (*B. officinalis*; Ref. 29), and *C. elegans* (30) Δ-6 desaturases reveals a high level of homology with a comparable domain within mammalian cytochrome b_5 (*Cyto.b5*) (31). Amino acids that are identical between the Δ-6 desaturases and cytochrome b_5 are highlighted in *black*; amino acids that are highly homologous between the desaturases and cytochrome b_5 are highlighted in *gray*. *Asterisk*, two heme-binding histidines found in cytochrome b_5 (31). *B*, the three histidine-rich regions conserved in membrane desaturases. The histidines within these regions are highlighted in *black*. The amino acids that are identical in all the Δ-6 desaturases listed are highlighted in *dark gray*.

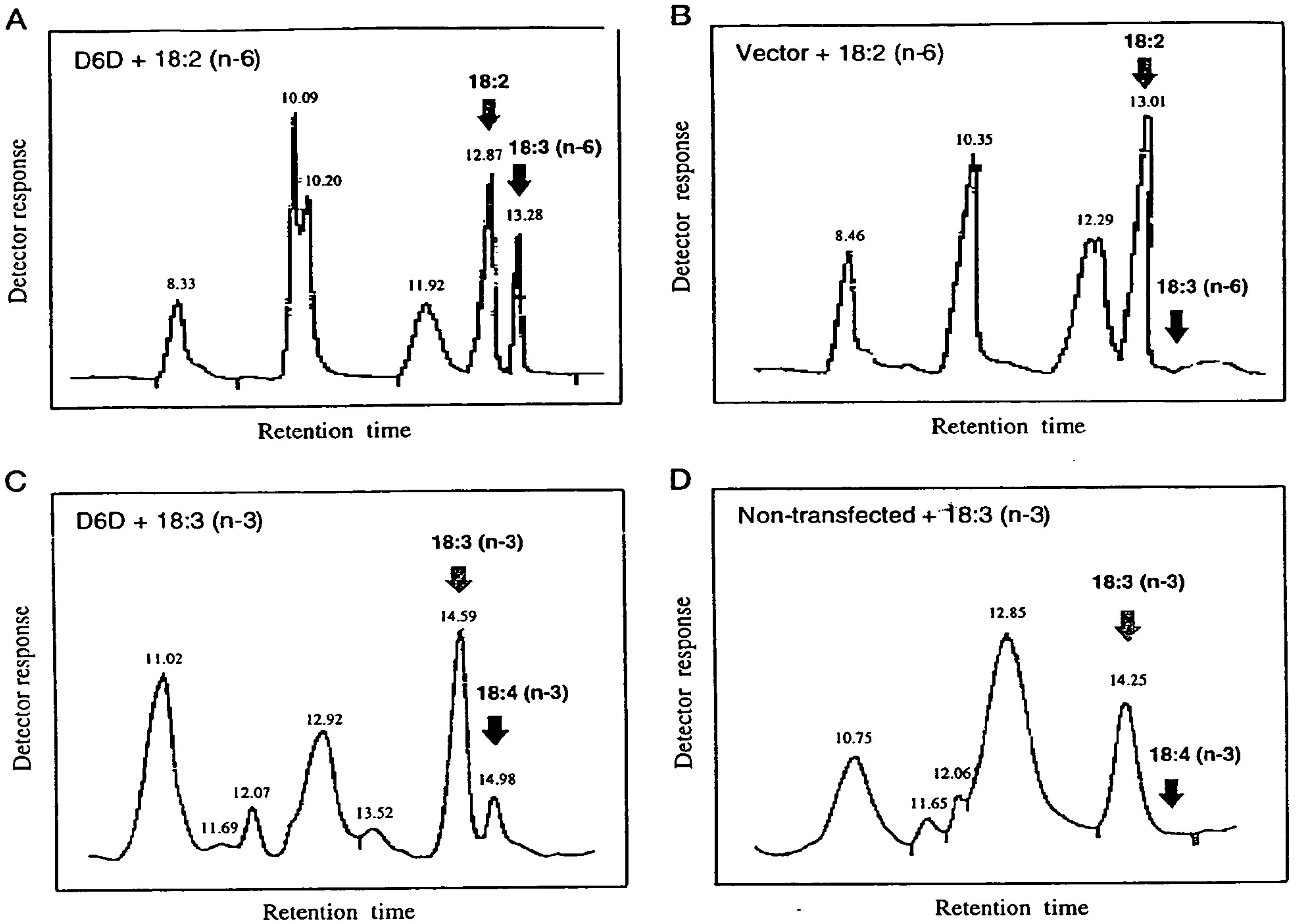


FIG. 4. **Expression of mouse Δ-6 desaturase in rat hepatocytes and CHO cells.** *A* shows the conversion of 18:2(n-6) to 18:3(n-6) by hepatocytes that were transfected with the pcDNA3.1 vector containing the mouse Δ-6 desaturase (*D6D*) ORF. When hepatocytes were transfected with pcDNA3.1 vector lacking the Δ-6 desaturase ORF, there was no detectable conversion of 18:2(n-6) to 18:3(n-6) (*B*). The media of CHO cells incubated with albumin-bound 18:3(n-3) and transfected with the Δ-6 desaturase expression vector pcDNA3.1 contained 18:4(n-3) (*C*), whereas the media of nontransfected CHO cells contained no detectable 18:4(n-3) (*D*). Retention times for the fatty acids are shown *above* the respective peaks. The identity of each peak was confirmed using individual fatty acid methyl ester standards.

Northern analysis of Δ-6 desaturase expression revealed that human Δ-6 desaturase mRNA is a single transcript of approximately 3.2 kb and is expressed in a wide array of tissues including the brain, liver, lung, and heart (Fig. 5*C*). The level of Δ-6 desaturase mRNA in the liver was approximately the same as that found in the lung and heart, but the abundance of Δ-6 desaturase in the human brain was severalfold higher (Fig. 5*C*). In addition to the tissues examined by Northern analysis, a search of the EST database revealed that Δ-6 desaturase mRNA is expressed in the human fetus and fetal heart as well as in the 13-day-old mouse embryo heart.

DISCUSSION

The purification and characterization of mammalian Δ-6 desaturase have been difficult because of its instability. In fact, there has been only one report, in 1981, that describes the purification of a putative linoleoyl-CoA desaturase from rat liver (33). Because of the problems encountered in the purification of Δ-6 desaturase, we have used the EST database to clone and characterize the mouse and human Δ-6 desaturase enzyme. Interestingly, a comparison of the rat liver linoleoyl-CoA desaturase with the Δ-6 desaturase peptide predicted by the ORF of both the mouse and human cDNAs indicates that the two proteins are markedly different. First, the ORF for mouse and human Δ-6 desaturase predicts a protein that is 52.2 kDa, whereas the size of the linoleoyl-CoA desaturase was cited to be 66 kDa (33). Second, the nucleotide sequence of the mouse and human Δ-6 desaturase ORFs predicts that these peptides contain 30 histidines (Fig. 1*B*). Moreover, many of these histidines are organized into distinct histidine-rich domains. Such domains are characteristic of all membrane-associated desaturases (32). In contrast, the reported amino acid composition of linoleoyl-CoA desaturase indicates that it contains only 15 histidine residues (33). Unfortunately, sequence information for linoleoyl-CoA desaturase is not available, because the purification of linoleoyl-CoA desaturase has never been replicated since the initial report. Clearly, the Δ-6 desaturase and the putative linoleoyl-CoA desaturase are distinctly different proteins. It is possible the liver contains two Δ-6 desaturase enzymes. In fact, metabolic studies suggest that there may be two isoforms of Δ-6 desaturase (34, 35): (*a*) one that catalyzes the initial desaturation of 18:2(n-6) or 18:3(n-3), and (*b*) another that catalyzes the conversion of 24:5(n-3) to 24:6(n-3). The cloning of the Δ-6 desaturase should now permit us to determine whether isoforms of Δ-6 desaturase do exist.

In addition to the histidine-rich domains, the mammalian Δ-6 desaturase contains a distinct cytochrome b_5-like domain

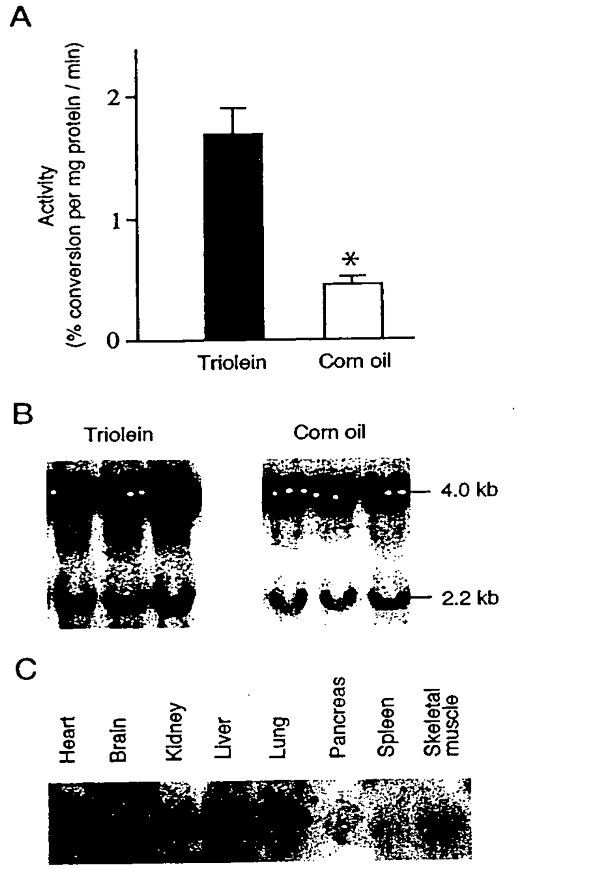


FIG. 5. **Nutritional regulation and tissue distribution of mammalian Δ-6 desaturase.** Mice were fed a high-glucose diet containing 10% corn oil or 10% triolein. Hepatic Δ-6 desaturase activity, which is expressed in *A* as means ± S.E., was significantly lower in mice fed corn oil than in mice fed triolein ($p < 0.001$). The abundance of hepatic Δ-6 desaturase mRNA was determined by Northern analysis (*B*). The abundance of the 4.0- and 2.2-kb Δ-6 desaturase transcripts was quantified by radio-imaging. The cpm of the ^{32}P-labeled probe associated with the 4.0- and 2.2-kb transcripts was 2509 ± 154 and 327 ± 17; it was 1264 ± 66 and 185 ± 9 for the triolein and corn oil groups, respectively ($p <$ [illegible]3). *C* depicts the abundance of Δ-6 desaturase mRNA found in a [illegible]ty of adult male human tissues. Each lane contains 20 μg of total RNA. Unlike mice, only one Δ-6 desaturase transcript with an approximate size of 3.2 kb was detected in human tissues. Comparable results were obtained with three different Northern blots.

that is also characteristic of plant (*Borage*) and *C. elegans* Δ-6 desaturases (29, 30) but is not a component of the mammalian Δ-9 desaturase (36). Early reconstitution studies with Δ-9 desaturase indicated that the conversion of 18:0(*n*-9) to 18:1(*n*-9) required Δ-9 desaturase, cytochrome b_5 reductase, and cytochrome b_5 itself (36). It has been assumed from these early studies that all mammalian desaturases require cytochrome b_5 for enzymatic activity (16, 36). However, the cytochrome b_5-like domain of yeast OLE1 was recently reported to replace the requirement for cytochrome b_5; *i.e.* desaturation occurred in the absence of cytochrome b_5, and removal of the cytochrome b_5-like domain rendered the OLE1 enzyme inactive (37). This observation raises the possibility that cytochrome b_5 reductase transfers electrons to the catalytic domain of the Δ-6 desaturase via the cytochrome b_5-like domain, and not via cytochrome b_5 per se.

Hepatic Δ-6 desaturase enzymatic activity varies with hormonal and nutritional manipulation (15, 16, 20, 38). For example, insulin deficiency and fasting reduce Δ-6 desaturase enzymatic activity, whereas the administration of insulin or refeeding increases its activity (39). In addition to being affected by fasting and feeding, hepatic Δ-6 desaturase enzymatic activity is highly dependent upon the composition of dietary fat (16). Specifically, the ingestion of fats that are low in essential fatty acids (*e.g.* butter) results in higher levels of enzyme activity than the consumption of fats (*e.g.* corn oil) that are rich in essential fatty acids (16). Northern analysis indicates that the increase in hepatic Δ-6 desaturase activity associated with the consumption of an essential fatty acid-deficient diet is paralleled by a comparable increase in the hepatic abundance of Δ-6 desaturase mRNA (Fig. 5). Thus, it appears that the activity of hepatic Δ-6 desaturase is largely regulated by pretranslational events. However, this may not be the case in all tissues. Specifically, Δ-6 desaturase activity is reportedly very low in nonhepatic tissues (16–18). Because of this low enzymatic activity in nonhepatic tissues, the liver has been considered to be the primary site of 20:4(*n*-6), 20:5(*n*-3), and 22:6(*n*-3) production for peripheral tissue utilization (17). However, Northern analysis of RNA from a number of different human tissues challenges this concept (Fig. 5*C*). For example, the level of Δ-6 desaturase mRNA in the human liver was comparable to that found in the human lung and heart. Moreover, the abundance of Δ-6 desaturase mRNA in the adult human brain was severalfold greater than that in the human liver (Fig. 5*C*). This high level of expression is certainly very consistent with the fact that >30% of the human brain phospholipid consists of 20- and 22-carbon polyenoic fatty acids (5, 6, 40). However, such high expression is in conflict with the reports that brain microsomes have a rate of Δ-6 desaturation that is only 10–15% of that found in the liver (18, 41). These data suggest that Δ-6 desaturase enzymatic activity may be determined by tissue-specific mechanisms that involve both pre- and post-translational events.

In conclusion, Δ-6 desaturase catalyzes the rate-limiting step in the conversion of 18:2(*n*-6) and 18:3(*n*-3) to the long chain polyenoic fatty acids 20:4(*n*-6) and 20:5(*n*-3) and 22:6(*n*-3), respectively (15). These long chain polyenoic fatty acids are essential for a large number of biological functions including inflammatory responses (4), brain development (2), retina function and cognition (1, 3), signal transduction (42, 43), reproduction (4), fetal growth (9), cell differentiation (14), and gene regulation (11–13). Not surprisingly, physiological conditions that are associated with low levels of Δ-6 desaturase activity may have a pronounced impact on a wide array of biological functions. For example, an impaired conversion of 18:2(*n*-6) to 18:3(*n*-6) appears to be associated with reduced nerve conductivity in human diabetics (7). Similarly, the low rate of 18:3(*n*-3) conversion to 20:5(*n*-3) and 22:6(*n*-3) observed in newborn infants is highly correlated with impaired retina function and reduced cognitive development (1). Now that the Δ-6 desaturase has been cloned, we can begin to define the role that Δ-6 desaturation may play in an apparently wide array of physiological processes.

REFERENCES

1. Birch, E. E., Hoffman, D. R., Uauy, R., Birch. D. G., and Prestidge, C. (1998) *Pediatr. Res.* **44**, 201–209
2. Uauy, R., Birch, E., Birch, D., and Peirano, P. (1992) *J. Pediatr.* **120**, S168–S180
3. Neuringer, M., Reisbick, S., and Janowsky, J. (1994) *J. Pediatr.* **125**, S39–S47
4. Samuelsson, B. (1983) *Biosci. Rep.* **3**, 791–813
5. Martinez, M. (1992) *J. Pediatr.* **120**, S129–S138
6. Crawford, M. A., Costeloe, K., Ghebremeskel, K., Phylactos, A., Skirvin, L., and Stacey, F. (1997) *Am. J. Clin. Nutr.* **66**, 1032S–1041S
7. Horrobin, D. F. (1997) *Diabetes* **46**, S90–S93
8. Lands, W. E. M. (1991) *Annu. Rev. Nutr.* **11**, 41–60
9. Carlson, S. E., Werkman, S. H., Peeples, J. M., Cooke, R. J., and Tolley, E. A. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 1073–1077
10. Werkman, S. H., and Carlson, S. E. (1996) *Lipids* **31**, 91–97
11. Clarke, S. D., and Jump, D. B. (1994) *Annu. Rev. Nutr.* **14**, 83–98
12. Clarke, S. D., Baillie, R., Jump, D., and Nakamura, M. T. (1997) *Ann. N. Y. Acad. Sci.* **827**, 178–187
13. Forman, B. M., Chen, J., and Evans, R. M. (1997) *Proc. Natl. Acad. Sci. U. S. A.*

94, 4312–4317
14. Spiegelman, B. M., and Flier, J. S. (1996) *Cell* **87**, 377–389
15. Sprecher, H. (1981) *Prog. Lipid Res.* **20**, 13–22
16. Brenner, R. R. (1989) in *The Role of Fats in Human Nutrition* (Vergroesen, A. J., and Crawford, M., eds), 2nd Ed., pp. 45–79, Academic Press, San Diego, CA
17. Scott, B. L., and Bazan, N. G. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2903–2907
18. Bourre, J. M., Piciotti, M., and Dumont, O. (1990) *Lipids* **25**, 354–356
19. Melin, T., and Nilsson, A. (1997) *Prostaglandins Leukotrienes Essent. Fatty Acids* **56**, 437–442
20. Peluffo, R. O., Nervi, A. M., and Brenner, R. R. (1976) *Biochim. Biophys. Acta* **441**, 25–31
21. Kawashima, Y., Musoh, K., and Kozuka, H. (1990) *J. Biol. Chem.* **265**, 9170–9175
22. Igal, R. A., Mandon, E. C., and de Gomez Dumm, I. N. (1991) *Mol. Cell. Endocrinol.* **77**, 217–227
23. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
24. Kozak, M. (1981) *Nucleic Acids Res.* **9**, 5233–5252
25. Armstrong, M. K., Blake, W. L., and Clarke, S. D. (1991) *Biochem. Biophys. Res. Commun.* **177**, 1056–1061
26. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
27. Reddy, A. S., Nuccio, M. L., Gross, L. M., and Thomas, T. L. (1993) *Plant Mol. Biol.* **27**, 293–300
28. Ntambi, J. M., Buhrow, S. A., Kaestner, K. H., Christy, R. J., Sibley, E., Kelly, T. J., Jr., and Lane, M. D. (1988) *J. Biol. Chem.* **263**, 17291–17300
29. Sayanova, O., Smith, M. A., Lapinskas, P., Stobart, A. K., Dobson, G., Christie, W. W., Shewry, P. R., and Napier, J. A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 4211–4216
30. Napier, J. A., Hey, S. J., Lacey, D. J., and Shewry, P. R. (1998) *Biochem. J.* **330**, 611–614
31. Lederer, F. (1994) *Biochimie (Paris)* **76**, 674–692
32. Shanklin, J., Whittle, E., and Fox, B. G. (1994) *Biochemistry* **33**, 12787–12794
33. Okayasu, T., Nagao, M., Ishibashi, T., and Imai, Y. (1981) *Arch. Biochem. Biophys.* **206**, 21–28
34. Voss, A., Reinhart, M., Sankarappa, S., and Sprecher, H. (1991) *J. Biol. Chem.* **266**, 19995–20000
35. Marzo, I., Alava, M. A., Pineiro, A., and Naval, J. (1996) *Biochim. Biophys. Acta* **1301**, 263–272
36. Strittmatter, P., Spatz, L., Corcoran, D., Rogers, M. J., Setlow, B., and Redline, R. (1974) *J. Biol. Chem.* **71**, 4565–4569
37. Mitchell, A. G., and Martin, C. E. (1995) *J. Biol. Chem.* **270**, 29766–29772
38. Nakamura, M. T., Phinney, S. D., Tang, A. B., Oberbauer, A. M., German, J. B., and Murray, J. D. (1996) *Lipids* **31**, 139–143
39. Poisson, J.-P., and Cunnane, S. C. (1991) *J. Nutr. Biochem.* **2**, 60–70
40. Uauy, R., Peirano, P., Hoffman, D., Mena, P., Birch, D., and Birch, E. (1996) *Lipids* **31**, S167–S176
41. Clandinin, M. T., Wong, K., and Hacker, R. R. (1985) *Biochem. J.* **226**, 305–309
42. Wolf, B. A., Turk, J., Sherman, W. R., and McDaniel, M. L. (1986) *J. Biol. Chem.* **261**, 3501–3511
43. Piomelli, D. (1993) *Curr. Opin. Cell Biol.* **5**, 274–280

(7)

Plant Molecular Biology

An International Journal on Fundamental Research and Genetic Engineering

Volume 22 / 1993

Kluwer Academic Publishers
Dordrecht / Boston / London

Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 4211–4216, April 1997
Plant Biology

Expression of a borage desaturase cDNA containing an N-terminal cytochrome b_5 domain results in the accumulation of high levels of Δ^6-desaturated fatty acids in transgenic tobacco

OLGA SAYANOVA*†, MARK A. SMITH*†, PETER LAPINSKAS‡, A. KEITH STOBART†, GARY DOBSON§, WILLIAM W. CHRISTIE§, PETER R. SHEWRY*, AND JOHNATHAN A. NAPIER*¶

*Institute of Arable Crops Research–Long Ashton Research Station, University of Bristol, Long Ashton, Bristol BS18 9AF, United Kingdom; †Department of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, United Kingdom; ‡Scotia Pharmaceuticals Ltd., Weyvern House, Peasmarsh, Guildford, Surrey GU3 1NA, United Kingdom; and §Department of Chemistry, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland

Communicated by Jake MacMillan, University of Bristol, United Kingdom, January 23, 1997 (received for review November 28, 1996)

ABSTRACT **γ-Linolenic acid (GLA; C18:3 $\Delta^{6,9,12}$) is a component of the seed oils of evening primrose (*Oenothera* spp.), borage (*Borago officinalis* L.), and some other plants. It is widely used as a dietary supplement and for treatment of various medical conditions. GLA is synthesized by a Δ^6-fatty acid desaturase using linoleic acid (C18:2 $\Delta^{9,12}$) as a substrate. To enable the production of GLA in conventional oilseeds, we have isolated a cDNA encoding the Δ^6-fatty acid desaturase from developing seeds of borage and confirmed its function by expression in transgenic tobacco plants. Analysis of leaf lipids from a transformed plant demonstrated the accumulation of GLA and octadecatetraenoic acid (C18:4 $\Delta^{6,9,12,15}$) to levels of 13.2% and 9.6% of the total fatty acids, respectively. The borage Δ^6-fatty acid desaturase differs from other desaturase enzymes, characterized from higher plants previously, by the presence of an N-terminal domain related to cytochrome b_5.**

Δ^6-Desaturated fatty acids are of major importance in animal cells as they have roles in the maintenance of membrane structure and function, in the regulation of cholesterol synthesis and transport, in the prevention of water loss from the skin, and as precursors of eicosanoids, including prostaglandins and leucotrienes (1). In animals, members of this class of fatty acids are synthesized from the essential fatty acid linoleic acid (C18:2 $\Delta^{9,12}$), the first step being the desaturation to γ-linolenic acid (GLA; C18:3 $\Delta^{6,9,12}$) catalyzed by a Δ^6-desaturase (1). Decreased activity of this key enzyme, observed for example in aging, stress, diabetes, eczema, and some infections, or increased catabolism of GLA resulting from oxidation or more rapid cell division (e.g., in cancer or inflammation) may lead to a deficiency of GLA (reviewed in ref. 2). Clinical trials have shown that dietary supplementation with GLA may be effective in treating a number of such conditions (e.g., atopic eczema, mastalgia, diabetic neuropathy, viral infections, and some types of cancer; ref. 2). Oils containing GLA are therefore widely used as a general health supplement and have been registered for pharmaceutical use.

In the plant kingdom, GLA is an uncommon fatty acid (3). Only a small number of higher plant species synthesize GLA, and in many of these, the fatty acid is found exclusively in the seed. GLA is also present in some fungi (e.g., *Mucor javanicus*) and cyanobacteria (3). Major commercial sources of GLA (4) are evening primrose (*Oenothera* spp.), in which GLA accounts for about 8–10% of the seed oil and borage (starflower) (*Borago officinalis* L.) seeds that contain some 20–25% GLA. These plants, however, suffer from poor agronomic performance and low yield; borage, for example, produces 300–600 kg/ha in the United Kingdom (4) compared with about 3 t/ha for oilseed rape. There is therefore considerable interest in both increasing the GLA content of existing crops and the production of GLA in a conventional oil crop (such as high linoleate rape).

In the higher plant cell, the synthesis of saturated fatty acids with chain lengths up to C18 and monounsaturated fatty acids (generally with a double bond at the Δ^9 position) occurs in the plastid. Further desaturation can then occur either in the plastid or on the endoplasmic reticulum (ER; ref. 5). The desaturase enzymes of the plastid require reduced ferredoxin as an electron donor and are either soluble enzymes acting on saturated acyl-ACP substrates or membrane-bound enzymes using unsaturated fatty acids esterified to complex lipids such as monogalactosyldialglycerol. In contrast, the ER-located Δ^{12}- and Δ^{15}-desaturases use fatty acids located at the *sn*-2 position of phosphatidylcholine as substrates, and cytochrome b_5 as a cofactor (5, 6). The Δ^6-fatty acid desaturase in the developing cotyledons of borage is similar to the Δ^{12}- and Δ^{15}-desaturases in its location and substrate specificity (oleate/linoleate at the *sn*-2 position of phosphatidylcholine), and is assumed to use cytochrome b_5 as its electron donor (7, 8). In addition, α-linolenic acid esterified to phosphatidylcholine may act as a substrate, resulting in the accumulation of octadecatetraenoic acid (OTA; C18:4 $\Delta^{6,9,12,15}$) in borage leaves (9).

We describe the isolation of a cDNA clone encoding the Δ^6-fatty acid desaturase from developing seeds of borage, using a PCR-based strategy. The identity of the cDNA has been confirmed by functional expression and analysis in transgenic tobacco plants. The encoded protein differs from other membrane-bound fatty acid desaturases of plants, such as those encoded by the FAD genes of *Arabidopsis* (10, 11), in that the desaturase domain is preceded at the N terminus by a sequence that is related to cytochrome b_5 (12), the haemprotein involved in electron transport to other ER-located fatty acid desaturases ($\Delta^{12,15}$) from higher plants (8, 13).

MATERIALS AND METHODS

Nucleic Acid Manipulations. Total RNA was isolated from developing seeds of borage (*B. officinalis*) using guanidinium thiocyanate according to the method described in ref. 14. Poly(A)$^+$ RNA was purified from total RNA using oligo(dT) cellulose according to standard methods (15) and was used as

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: GLA, γ-linolenic acid; OTA, octadecatetraenoic acid; ER, endoplasmic reticulum; FAMe, fatty acid methyl ester; DMOX, 4,4-dimethyloxazoline; MS, mass spectrometry.
Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. U79010 and U79011).
¶To whom reprint requests should be addressed.

a template for cDNA library construction. Single-stranded cDNA was synthesized from total RNA using the Reverse Transcription System (Promega) according to the supplier's instructions and used as a template for PCR amplification with degenerate primers. All nucleotide sequences were determined by the dideoxy chain termination method (15), and aligned using the GCG 8 program (16).

PCR-Based Cloning. Two highly degenerate primers were synthesized for cDNA screening: forward primer A, 5′-GCGAATTC(A/G)TXGGXCA(T/C)GA(T/C)TG(T/C)G-GXCA-3′ (fully degenerate to the conserved amino acid sequence GHDCGH), and reverse primer B, 5′-GCGAATT-CATXT(G/T)XGG(A/G)AAXA(G/A)(A/G)TG(A/G)-TG-3′ (fully degenerate to conserved amino acid sequence HHLFP), where X substitutes nucleotides AGTC. Each primer contained an *Eco*RI site (underlined) at the 5′ end to facilitate subsequent manipulations. These primers were used for PCR amplification with cDNA transcribed from total RNA. Reactions were run on a Perkin–Elmer Cetus DNA thermal cycler using a program of 1 min at 94°C, 1 min at 45°C, and 2 min at 72°C for 35 cycles followed by extension for 10 min at 72°C. PCR amplification products were separated on 1.0–2% agarose gels. PCR fragments of the expected length (600–700 bp) were purified using the Wizard DNA purification system (Promega), ligated into pGEM-T Vector according to the pGEM-T Vector Cloning Kit (Promega), and transformed into XL1-blue *Escherichia coli* cells. Plasmid DNA was purified and sequenced using the Promega miniprep system.

Library Screening. Poly(A)$^+$ mRNA from developing seeds of borage was used as the template for the synthesis of a cDNA library; custom synthesis and packaging being carried out by CLONTECH. The cDNA was inserted into the *Eco*RI site of the phage vector λ ZAPII, and the resultant DNA was packaged into phage particles. The cDNA bank contained 2.0×10^6 clones with an average insert size of 2.0 kb. Filter replicas of this library were hybridized with the labeled DNA probe pBdes1 and with a tobacco cDNA encoding cytochrome b_5 (17). Radiolabeling of DNA and screening of phage libraries were conducted using standard techniques (15). The full-length cDNA clone pBdes6 was isolated and sequenced on both strands.

Northern Blot Analysis. RNA was separated by electrophoresis through 1% formaldehyde agarose gel, transferred to nylon membrane (Hybond N, Amersham), and bound by exposure to UV light for 2 min. Probes were made from the cDNA clone pBdes6 by random priming (15). The filters were hybridized and washed as described in ref. 17 and then exposed to x-ray film at −80°C using an intensifying screen.

Plant Transformation. To facilitate preparation of plant expression constructs, flanking *Sal*I and *Sma*I restriction enzyme sites were added to the coding region of clone pBdes6 by PCR amplification. Two oligonucleotides were synthesized based on the pBdes6 coding sequence: primer C, 5′-GCGTCGAC**ATG**GCTGCTCAAATCAAG-3′ (annealing to the initiating methionine, indicated in boldface type), and primer D, 5′-GCCCGGG**TTA**ACCATGAGTGTGAAG-3′ (annealing up to the complement of the stop codon, indicated in boldface type). The *Sal*I (primer C) and *Sma*I (primer D) restriction sites are underlined. The PCR product was purified and subcloned into the vector pJD330 (18) to generate the plasmid p35Bdes6. Digestion of p35Bdes6 with *Xba*I released fragment of ≈2,200 bp containing the ORF of the borage pBdes6, together with regulatory elements consisting of the cauliflower mosaic virus 35S promoter, an Ω-translational enhancer from tobacco mosaic virus (19) and the nopaline synthase (*nos*) termination sequence. This *Xba*I fragment was gel purified and cloned into pBIN19 (20) to obtain the plasmid pNTdes6, which was transformed into *Agrobacterium tumefaciens* strain LBA4404 by electroporation. Tobacco (*Nicotiana tabacum* cv. NVS) was transformed with the plant expression plasmid according to standard procedures (21). Initial transformants were selected on 50 μg/ml kanamycin and then transferred to 100 μg/ml kanamycin. Plants were maintained in axionic culture under controlled conditions.

Fatty Acid Analysis. Lipids were extracted from leaves of transformed and control tobacco plants by homogenization in MeOH-$CHCl_3$ using a modification of the method of Bligh and Dyer (22). The resulting $CHCl_3$ phase was evaporated to dryness under nitrogen gas, and the samples were transmethylated with 1 M HCl in methanol at 80°C for 1 h. Fatty acid methyl esters (FAMes) were extracted in hexane and purified using a small column packed with Florisil. Analysis of FAMes was conducted using a Hewlett Packard 5880A Series Gas Chromatograph equipped with a 25 M × 0.32 mm RSL-500BP bonded capillary column and a flame ionization detector. Fatty acids were identified by comparison of retention times with FAMe standards (Sigma) separated on the same GC. Quantitation was carried out using peak height area integrals expressed as a total of all integrals.

GC–Mass Spectrometry (MS) Analysis. Fatty acid 4,4-Dimethyloxazoline (DMOX) derivatives were prepared for GC-MS analysis by a modification of the method of Fay and Richli (23). Lipid samples (extracted from tobacco leaves as described above) were heated at 180°C in 2-amino-2-methyl-1-propanol under N_2 for 18 h. After cooling to room temperature dichloromethane and water were added. The DMOX derivatives were recovered in the dichloromethane, passed through a column of anhydrous sodium sulfate to remove water, and dried under a stream of N_2. To remove any contaminating polar material, the samples were taken up in hexane, passed through a short Florisil column, and evaporated to dryness. The samples were then dissolved in an appropriate volume of hexane for GC-MS analysis. Fatty acid DMOX derivatives were analyzed by GC-MS on a Hewlett Packard 5890 Series II Plus gas chromatograph equipped with a 50 M × 0.25 mm BPX70™ capillary column connected directly to a Hewlett Packard 5989B MS Engine quadropole mass spectrometer operating at an ionization energy of 70 eV and emission current of 300 μA. Mass spectra were interpreted by comparison to the mass spectra of DMOX derivatives of GLA and OTA prepared from blackcurrant oil, which is known to contain both these fatty acids (24), using the interpretation rules of ref. 25.

RESULTS

PCR-Based Cloning of Membrane-Bound Desaturases. Comparisons of the deduced amino acid sequences of membrane-bound fatty acid desaturases (and related proteins) from mammals, fungi, insects, higher plants, and cyanobacteria reveal three highly conserved regions (boxes) containing histidine residues (26). Since the borage seed Δ^6-desaturase is membrane-bound (27), two highly degenerate primers were constructed based on the sequences of the first and third histidine boxes present in the membrane-bound Δ^{12}- and Δ^{15}-fatty acid desaturases of plants. These primers were used in PCRs with cDNA transcribed from total RNA of developing cotyledons of borage. PCR products of the predicted length (600–700 bp) were cloned and sequenced, allowing them to be classified into three groups: 45% showed similarity to other proteins (i.e., not fatty acid desaturases), 35% resembled Δ^{12}-desaturases, and 20% formed a separate group that showed some similarity to both Δ^{12}- and Δ^{15}-desaturases but was clearly distinct from the second group.

Sequencing of a representative clone (pBdes1) from the third group revealed an ORF of 228 aa with three putative histidine boxes. Alignment of the deduced amino acid sequence with those of known desaturases (data not shown) showed highest similarity to the Δ^{15}-desaturases, although the actual level of identity was low (less than 30%). Since borage

seed oil contains little or no α-linolenic acid, it is unlikely that high levels of transcripts for Δ^{15}-desaturases would be present in the developing seeds. It was therefore considered likely that the pBdes1 PCR product encoded part of a putative Δ^{6}-desaturase.

To isolate a full-length clone corresponding to the pBdes1 PCR product, the insert was used to probe a borage developing seed cDNA library constructed in λ ZAPII. A total of 3×10^5 plaques were screened, and 20 individual phage clones that hybridized with the pBdes1 DNA probe were identified and purified by further rounds of hybridization. Restriction enzyme digestion of 15 clones recovered from positive plaques showed the presence of single inserts that hybridized with the probe, ranging from 700 to 1,800 bp in length. One of these, termed pBdes6, containing an insert of 1,800 bp, was chosen for detailed analysis.

pBdes6 encodes a 1,344-bp ORF, preceded by a 41-bp 5′ untranslated region. The coding region was followed by a stop codon and a 345-bp untranslated region with a poly(A) tail. The ORF encoded 448 aa, corresponding to a putative protein with an M_r of about 50,000, which is significantly larger than the predicted M_r of other microsomal desaturases such as the [illegible] and Δ^{15}-desaturases from *Arabidopsis* (refs. 10 and 11; Fig. 1). A degree of similarity to other fatty acid desaturases is clear, but only over a part of the coding sequence. The amino acid sequence from residues 144 to 448 showed about 17% identity with Δ^{15} (FAD3) (10) and Δ^{12} (FAD2) (10) desaturases from *Arabidopsis* and about 22% identity with a Δ^{6}-desaturase from the cyanobacterium *Synechocystis* (28). The whole sequence was also 60% identical to a cDNA clone of unknown function isolated from sunflower seeds (29). The three conserved histidine boxes that are characteristic of other membrane-bound desaturases were also present, and located at similar positions within the sequence. The distance between the first and second boxes was 32 aa, compared with 31 or 32 aa in Δ^{12}- and Δ^{15}-desaturases, and the distance between the second and third boxes was 172 aa, compared with 132–173 in other membrane-bound desaturases. The importance of these histidine boxes in catalysis has been demonstrated by site-directed mutagenesis of the soluble Δ^{9}-desaturase from rat and Δ^{12}-desaturase of *Synechocystis* (26, 30).

pBdes6 Encodes a Protein Containing a Cytochrome b_5-Like Heme-Binding Domain. The predicted hydrophobicity plot for the protein encoded by pBdes6 revealed a profile characteristic of a fatty acid desaturase, with the histidine boxes located in hydrophilic areas and separated by hydrophobic domains (not shown). The borage protein, however, contained a hydrophilic region at the N terminus longer than those of the membrane-bound Δ^{12}- and Δ^{15}-desaturases. Closer analysis showed significant sequence similarity between the first 90–100 aa at the N terminus of the protein encoded by pBdes6 and microsomal cytochrome b_5 proteins from higher plants (17). This similarity included the presence of seven of the eight invariant residues of the cytochrome b_5 class of proteins identified by Lederer (12). A heme-containing electron donor is required for fatty acid desaturation, and cytochrome b_5 is known to fulfill this function with membrane-bound fatty desaturases (Δ^{12} and Δ^{15}; refs. 8 and 13) and with the related Δ^{12}-hydroxylase (31). We therefore isolated a cDNA for cytochrome b_5 from the borage cDNA library using a tobacco cDNA (17) as a probe. Sequencing of this cDNA revealed an ORF encoding 132 aa which had some 80% sequence identity to cytochrome b_5 proteins from tobacco and rice (17). It also showed 32% sequence identity with the cytochrome b_5-related domain of the protein encoded by pBdes6 (Fig. 2). The identity is particularly high in regions previously identified as essential for cytochrome b_5 function, including the EHPGG motif in the heme-binding region.

Functional Analysis of pBdes6 in Transgenic Tobacco. To confirm the identity of pBdes6 as a Δ^{6}-fatty acid desaturase, the cDNA was transferred to tobacco plants under the control of an Ω-enhanced cauliflower mosaic virus 35S promoter via *Agrobacterium*-mediated gene transfer. Single leaves were removed from transformed and control plants, and FAMes were prepared from total lipid extracts and analyzed by GC (Fig. 3). Two peaks were observed in the chromatogram of

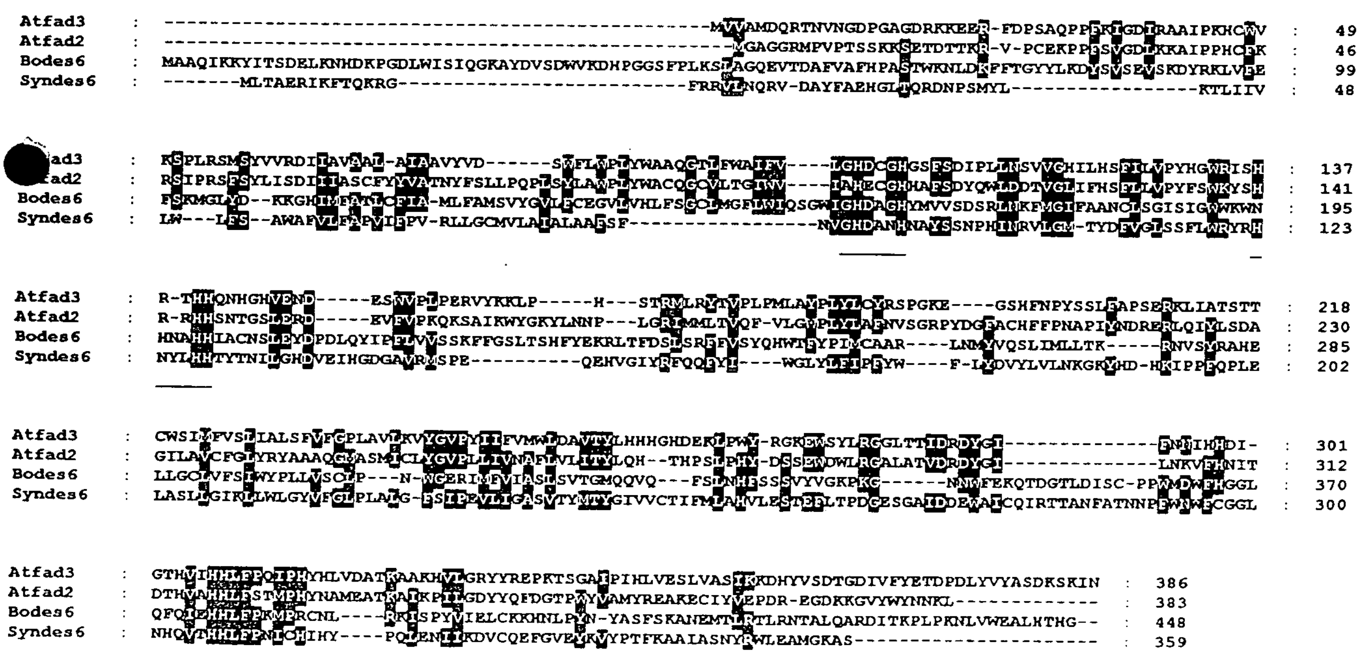


FIG. 1. Comparison of the deduced amino acid sequence of pBdes6 (labeled bodes6) with other desaturases. The entire coding sequence of pBdes6 was compared with the *Arabidopsis* FAD2 (atfad2) and FAD3 (atfad3) microsomal desaturases, as well as the cyanobacterial Δ^{6} desaturase (syndes6). Identical or conserved residues are boxed, and the conserved histidine boxes are underlined. The sequence of pBdes6 has been deposited in the GenBank database (accession no. U79010).

```
Bocytb5  :  ------MGKIFTLAEVAQHNNSKDCWLIINGKVYDVTKFLEDHPGGDDVLLSATGKDATDDFEDIGHSSSAKAM---LDEYYVGDIDSSSIPS---QVK  :   87
Ntcytb5  :  MDIMGGETKVFTLAEVSQHNNAKDCWLVISGKVYDVTKFLDDHPGGDEVLLSATGKDATDDFEDVGHSSSARAM---LDEYYVGDIDSATIPT---KTK  :   93
Oscytb5  :  --MSNDNKKVYTLEEVAKHNSKDDCWLIIGGKVYNVSKFLEDHPGGDDVLLSSTGKDATDDFEDVGHTTTARAM---MDEYYVGDIDTSTIPA---RTK  :   91
Bodes6   :  --MAAQIKKYITSDELKNHDKPGDLWISIQGKAYDVSDWVKDHPGGSFPLKSLAGQEVTDAFVAFHPASTWKNLDKFFTGYYLKDYSVSEVSKDYRKLV  :   97

Bocytb5  :  YTPPKQPLYNPD-----KTREFVIKLLQFLVPLVILAGAIGIRFYTKSSA-------------------------------------------------  :  132
Ntcytb5  :  YTPPNQPHYNQD-----KTSEFVVKLLQFLVPLIILGVAFGIRFYTKQSSA------------------------------------------------  :  139
Oscytb5  :  YVPPKQPHYNQD-----KTPEFIIKILQFLVPLAILGLAVAIRIYTKSESA------------------------------------------------  :  137
Bodes6   :  EEFSKMGLYDKKGHIMFATLCFIAMLFAMSVYGVLFCEGVLVHLFSGCLMGFLWIQSGWIGHDAGHYMVVSDSRLNKFMGIFAANCLSGISIGWWKWNH  :  196

Bocytb5  :  ----  :    -
Ntcytb5  :  ----  :    -
Oscytb5  :  ----  :    -
Bodes6   :  NAHH  :  200
```

FIG. 2. Comparison of the deduced amino acid sequence of pBdes6 with plant cytochrome b_5 sequences. The first 196 residues of pBdes6 (bodes6) were compared with cytochrome b_5 sequences from borage (bocytb5), rice (oscytb5), and tobacco (ntcytb5). The conserved heme-binding domain is underlined. The sequence of borage cytochrome b_5 has been deposited in the GenBank database (accession no. U79011).

FAMes from the transformants (Fig. 3*B*) not present in the control plants (Fig. 3*A*). These peaks had retention times identical to the FAMe standards of GLA (C18:3 $\Delta^{6,9,12}$) and OTA (C18:4 $\Delta^{6,9,12,15}$). Further analysis of the transformant by GC-MS analysis of fatty acid DMOX derivatives confirmed the identities of these peaks (Fig. 4). Both spectra contained abundant m/z 113 (McLafferty rearrangement ion) and 126 peaks typical of fatty acid DMOX derivatives (23). The spectrum of the putative GLA derivative (Fig. 4*A*) had a molecular ion at m/z 331, suggesting an octadecatrienoic fatty acid; gaps of 12 amu between m/z 194 and 206, and m/z 234 and 246, indicating double bonds at C9 and C12; and a prominent m/z 166/167 pair specific for a C6 double bond (25). The spectrum of the putative OTA (Fig. 4*B*) had an additional gap of 12 amu between m/z 274 and 286, indicating the presence of a C15 double bond.

The proportions of fatty acids in the total lipid fractions prepared from the leaves of the control and transformed tobacco plants are given in Table 1. GLA and OTA account for about 13% and 10% of the total, respectively, in the transgenic material and are absent in the control plants. The presence of both GLA and OTA indicates that the Δ^6-desaturase used both linoleic acid and α-linolenic acid as substrates, and this may be responsible for the decrease in α-linolenic acid observed in the transgenic line.

Northern Blot Analysis. To provide further evidence that the introduction of the borage cDNA into the tobacco genome was responsible for these novel desaturation products, total RNA was isolated from the leaves of either a GLA-positive transgenic tobacco plant or a control plant that had been subject to the same tissue culture regime. RNA was also isolated from developing borage seeds and leaves, and the samples were analyzed by Northern blotting and probed with the pBdes6 cDNA (Fig. 5). A positive hybridization signal of identical mobility was obtained from RNA isolated from borage seeds and the transgenic GLA-positive tobacco line, but not from the control tobacco plant. Prolonged exposure of the autoradiograph showed that low levels of the pBdes6 transcript (or related transcripts) were present in the RNA samples extracted from borage leaves, a result that is consistent with the observed accumulation of GLA in the leaves of this species (9).

DISCUSSION

We undertook to isolate a cDNA encoding a Δ^6-desaturase from borage using a degenerate PCR approach based on conserved amino acid sequence motifs in other microsomal fatty acid desaturases (26, 30). Previous studies (9, 27) had shown that the borage Δ^6-desaturase activity was associated with the microsomal membrane fraction and probably used cytochrome b_5 as an electron donor, like the microsomal Δ^{12} (FAD2) and Δ^{15} (FAD3) desaturases. The borage cDNA

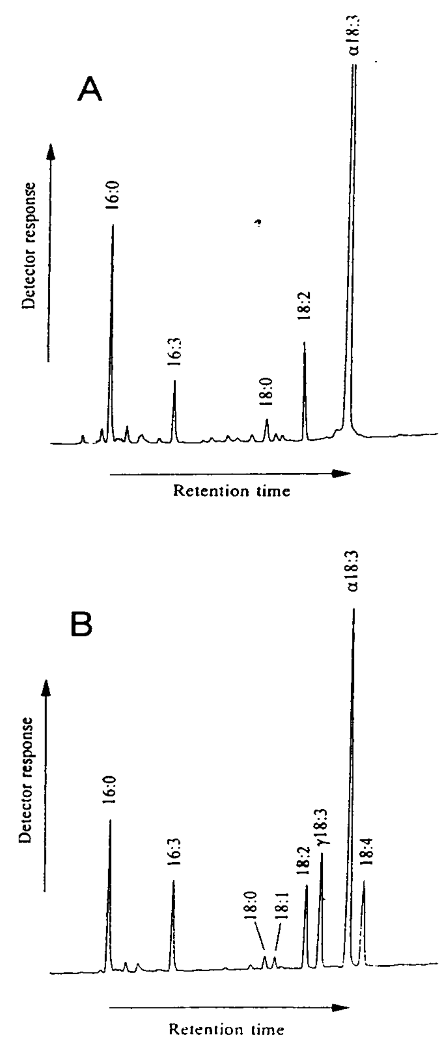


FIG. 3. Identification of GLA and OTA in transgenic tobacco by GC. Chromatograms of FAMes from leaf tissue of control tobacco plant (*A*) or plant transformed with pBdes6 (*B*). Two novel peaks are seen in *B*; these peaks have retention times identical to FAMe standards of GLA and OTA. The identity of peaks (as determined by comparison of retention times with those of known standards) is indicated. Detection was by flame ionization.

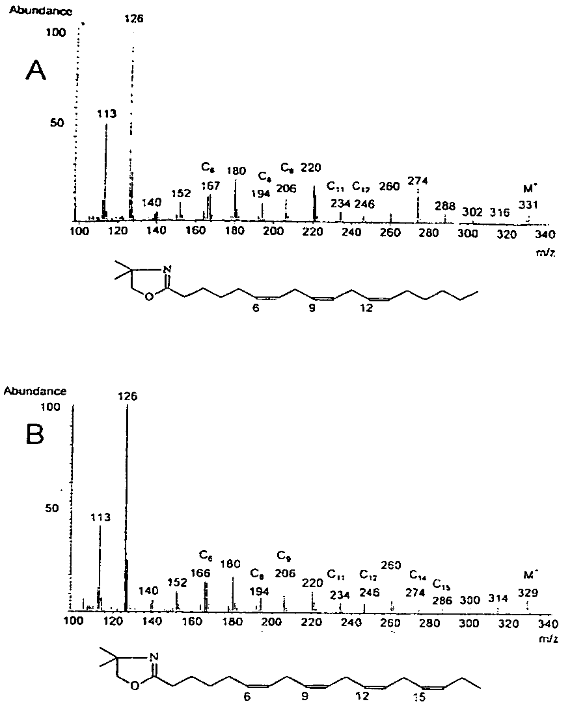


FIG. 4. Mass spectra of DMOX-derivatized fatty acids. Spectra of the fatty acids identified in Fig. 3 as GLA (*A*) and OTA (*B*). Details of the interpretation of the spectra are given in the text. The deduced structures of the fatty acid derivatives are shown.

clone (pBdes6) was confirmed to encode a Δ^6-desaturase by ectopic expression in the leaves of transgenic tobacco, resulting in the accumulation of the fatty acids GLA and OTA. The borage Δ^6-desaturase encoded by pBdes6 differed from previously characterized fatty acid desaturases from higher plants by the presence of an N-terminal extension related to the cytochrome b_5 class of heme-binding proteins (13). This domain is not present in the plant microsomal Δ^{12}- and Δ^{15}-desaturases (10, 11) or in the related Δ^{12}-hydroxylase (31) which have been cloned and functionally characterized in transgenic plants, although their use of microsomal cytochrome b_5 as an electron donor has been clearly demonstrated (8, 32). It is also clear that the N-terminal cytochrome b_5-related domain of pBdes6 is structurally distinct from the borage microsomal cytochrome b_5, as it does not contain the conserved hydrophobic C-terminal microsomal membrane anchor normally present in cytochrome b_5 proteins (17, 33). Since cytochrome b_5 usually functions in association with the

Table 1. Total fatty acid content of lipid extracts from leaves of a control tobacco plant and a plant transformed with the borage Δ^6 desaturase clone pBdes6

Acids		% Fatty acid	
		Control	Transformant
Palmitic	(C16:0)	16.3	14.0
Palmitoleic	(C16:1)	Trace	Trace
	(C16:3)	5.0	9.0
Stearic	(C18:0)	2.4	1.5
Oleic	(C18:1)	Trace	1.3
Linoleic	(C18:2)	9.1	9.5
γ-Linolenic	(C18:3)	ND	13.2
α-Linolenic	(C18:3)	65.1	40.1
OTA	(C18:4)	ND	9.6

Percentages were integrated from peak areas of GC traces shown in Fig. 3. ND, not detected.

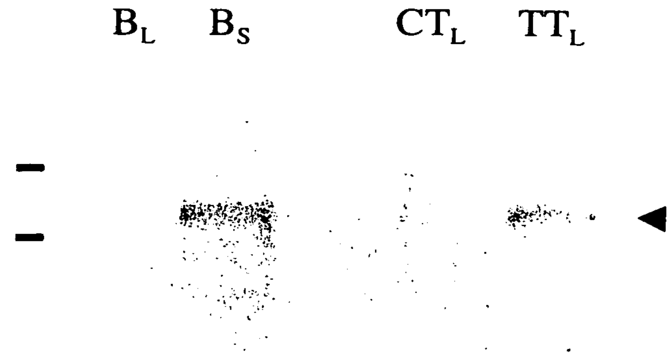


FIG. 5. Northern blot analysis of pBdes6 expression in borage and in transgenic tobacco. Total RNA (10 μg), extracted from borage leaves (B_L), borage seeds (B_S), control tobacco leaves (CT_L) or transgenic tobacco leaves (TT_L) was probed with ^{32}P-labeled pBdes6. After hybridization and high stringency washing, the resulting autoradiograph indicated expression of the pBdes6 transcript (≈2,000 bp; marked with the arrowhead) in borage seeds and transgenic tobacco leaves. The positions of the rRNA bands are indicated.

ER membrane, it is likely that the fusion protein described in this study has the same location and this is supported by the absence of any domains resembling chloroplast targeting transit sequences (34). Although the protein encoded by pBdes6 does not appear to have an N-terminal cleavable ER-targeting signal sequence (as judged by computer searching), the hydrophobic regions present in the protein would be sufficient to allow it associate with the endomembrane system. No obvious ER-retention motifs are present, but a potential glycosylation site is present at residues 278–280 (N-V-S).

Domains related to cytochrome b_5 are also present in a microsomal Δ^9-fatty acid desaturase (Ole1p, the OLE1 gene product) from yeast (35) and in other oxido/reductase enzymes (e.g., nitrate reductase, sulfite oxidase, and flavocytochrome b_2; ref. 12). In the yeast Δ^9-desaturase, this cytochrome b_5 domain exists as a 113-aa C-terminal fusion. Expression of OLE1 from a multicopy plasmid rescued yeast double mutants that lacked both OLE1 and microsomal cytochrome b_5 genes, unlike rescue by a rat microsomal Δ^9-desaturase, which required the presence of the cytochrome b_5 gene (35). Moreover, when the C-terminal b_5 domain of OLE1 was deleted, the yeast cells remained fatty acid auxotrophic, even in the presence of endogenous yeast cytochrome b_5, indicating that cytochrome b_5 is not able to act in trans to complement the loss of the cytochrome b_5 fusion domain of Ole1p (35). This suggests that the fusion domain plays an essential role in the desaturase reaction of this enzyme. A cDNA clone encoding a related cytochrome b_5 fusion protein has also been isolated from sunflower seeds (29), as noted above, but the corresponding protein has not been identified. In the sunflower protein, the cytochrome b_5 domain is fused to the N terminus of a putative desaturase sequence, as in the pBdes6 protein, and expression of this domain (≈120 residues) in *E. coli* (29) has shown that it is capable of undergoing reversible oxidation and reduction, indicating a functional heme group. Similar results were also obtained by expression of a tobacco cytochrome b_5 cDNA in *E. coli* (33). However, sunflower seeds do not accumulate GLA (3) and would therefore not be expected to possess an active Δ^6-desaturase. The substrate specificity of this sunflower protein is not known, and its role in fatty acid desaturation/hydroxylation-type reactions can only be inferred from sequence homology. Similarly, the functional and evolutionary significance of the existence of two types of membrane-bound desaturases in plants is not clear, although it can be suggested that the fusion of a cytochrome b_5 domain to the desaturase may facilitate a more efficient electron transfer. It is also

unclear why the yeast Ole1p desaturase has a C-terminal cytochrome b_5 domain, whereas the borage desaturase has an N-terminal cytochrome b_5 domain.

Recently, GLA and OTA accumulation in transgenic plants has been reported by Reddy and Thomas (36), who expressed a cyanobacterial Δ^6-desaturase gene in tobacco. The combined levels of GLA and OTA varied from about 2% to 4% of the leaf C:18 fatty acids, with only small differences depending on whether the protein was targeted to the plastid, cytoplasm, or ER lumen. This low level of activity is perhaps not surprising as the cyanobacterial Δ^6-desaturase differs from the ER-located higher plant desaturases in using ferredoxin rather than cytochrome b_5 as a cofactor. The cyanobacterial Δ^6-desaturase also resulted in the accumulation of comparatively higher levels of OTA than GLA, but the reason for this is not known. The levels of GLA and OTA accumulating in the leaves of transgenic tobacco plants expressing the borage desaturase encoded by pBdes6 account together for over 23% of total fatty acids, indicating the potential for producing GLA in transgenic oil crops. Sunflower would be particularly suitable in this respect as the presence of between 50% and 70% linoleic acid and with little or no α-linolenic acid (37) should facilitate the synthesis of high levels of GLA.

[illegible]R receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom. The work was supported by a grant from the Biotechnology and Biological Sciences Research Council under the Collaboration with Industry Scheme.

1. Horrobin, D. F. (1992) *Prog. Lipid Res.* **31**, 163–194.
2. Horrobin, D. F. (1990) *Rev. Contemp. Pharmacother.* **1**, 1–45.
3. Gunstone, F. D. (1992) *Prog. Lipid Res.* **31**, 145–161.
4. Fieldsend, A. F. (1995) *Biologist* **42**, 203–207.
5. Heinz, E. (1993) in *Lipid Metabolism in Plants*, ed. Moore, T. S. (CRC, Boca Raton, FL), pp. 34–89.
6. Stymne, S. & Stobart, A. K. (1993) in *Seed Storage Compounds*, eds. Shewry, P. R. & Stobart, A. K. (Oxford Univ. Press, Oxford), pp. 96–114.
7. Griffiths, G., Brechany, E. Y., Christie, W. W., Stymne, S. & Stobart, A. K. (1989) in *Biological Role of Plant Lipids*, eds. Biacs, P. A., Gruiz, K. & Kremmer, T. (Plenum, New York), pp. 151–154.
8. Smith, M. A., Cross, A. R., Jones, O. T. G., Griffiths, W. T., Stymne, S. & Stobart, A. K. (1990) *Biochem. J.* **272**, 23–29.
9. Griffiths, G., Brechany, E. Y., Jackson, F. M., Christie, W. W., Stymne, S. & Stobart, A. K. (1996) *Phytochemistry* **43**, 381–386.
10. Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H. & Somerville, C. R. (1992) *Science* **258**, 1353–1355.
11. Okuley, J., Lightner, J., Feldmann, K., Yadav, N., Lark, E. & Browse, J. (1994) *Plant Cell* **6**, 147–158.
12. Lederer, F. (1994) *Biochimie* **76**, 674–692.
13. Kerns, E. V., Hugly, S. & Somerville, C. R. (1991) *Arch. Biochem. Biophys.* **284**, 431–436.
14. Napier, J. A., Smith, M. A., Stobart, A. K. & Shewry, P. R. (1995) *Planta* **197**, 200–202.
15. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
16. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
17. Smith, M. A., Stobart, A. K., Shewry, P. R. & Napier, J. A. (1994) *Plant Mol. Biol.* **25**, 527–537.
18. Zakai, N., Ballas, N., Hershkovitz, M., Broido, S., Ram, R. & Loyter, A. (1993) *Plant Mol. Biol.* **21**, 823–830.
19. Gallie, D. R., Lucas, W. J. & Walbot, V. (1989) *Plant Cell* **1**, 301–311.
20. Bevan, M. (1984) *Nucleic Acid Res.* **12**, 8711–8722.
21. Draper, J., Scott, R., Armatige, P. & Walden, R. (1988) *Plant Genetic Transformation and Gene Expression: A Laboratory Manual* (Blackwell Scientific, Oxford).
22. Bligh, E. G. & Dyer, W. J. (1959) *Can. J. Biochem. Physiol.* **37**, 911–917.
23. Fay, L. & Richli, U. (1991) *J. Chromatogr.* **541**, 89–98.
24. Traitler, H., Willie, H. J. & Studer, A. (1988) *J. Am. Oil Chem. Soc.* **65**, 755–760.
25. Zhang, J. Y., Yu, Q. T., Liu, B. N. & Huang, Z. H. (1988) *Biomed. Environ. Mass Spectrom.* **15**, 33–44.
26. Shanklin, J., Whittle, E. & Fox, B. G. (1994) *Biochemistry* **33**, 12787–12794.
27. Stymne, S. & Stobart, A. K. (1986) *Biochem. J.* **24**, 385–393.
28. Reddy, A. S., Nuccio, M. L., Gross, L. M. & Thomas, T. L. (1993) *Plant Mol. Biol.* **22**, 293–300.
29. Sperling, P., Schmidt, H. & Heinz, E. (1995) *Eur. J. Biochem.* **232**, 798–805.
30. Avelange-Macherel, M.-H., Macherel, D., Wada, H. & Murata, N. (1995) *FEBS Lett.* **361**, 111–114.
31. van de Loo, F. J., Broun, P., Turner, S. & Somerville C. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 6743–6747.
32. Smith, M. A., Jonsson, L., Stymne, S. & Stobart, A. K. (1992) *Biochem. J.* **287**, 141–144.
33. Smith, M. A., Napier, J. A., Stymne, S., Tatham, A. S., Shewry, P. R. & Stobart, A. K. (1994) *Biochem. J.* **303**, 73–79.
34. Napier, J. A. (1995) in *Methods in Molecular Biology*, ed. Jones, H. (Humana, Totowa, NJ), Vol. 49, pp. 369–376.
35. Mitchell, A. G. & Martin, C. E. (1995) *J. Biol. Chem.* **270**, 29766–29772.
36. Reddy, A. V. & Thomas, T. L. (1996) *Nat. Biotechnol.* **14**, 639–642.
37. Salunkhe, D. K., Chavan, J. K., Adsule, R. N., Kadam, S. S. (1991) *World Oilseeds—Chemistry, Technology and Utilization* (Van Nostrand Reinhold, New York).

Chapter 3

Biosynthesis of γ-Linolenic Acid in the Cyanobacterium *Spirulina platensis*

Norio Murata, Patcharaporn Deshnium, and Yasushi Tasaka

Department of Regulation Biology,
National Institute for Basic Biology,
Myodaiji, Okazaki 444, Japan

Lipids and Fatty Acids in Cyanobacteria

Cyanobacteria are autotrophic prokaryotes with the capacity for photosynthesis. Cyanobacterial cells resemble the chloroplasts of plants in terms of both membrane structure and glycerolipid composition (1). There are three types of membrane in cyanobacterial cells: the plasma membrane, the outer membrane, and the thylakoid membrane. The thylakoid membranes are closed systems and are separate from the plasma membrane (1,2). This architecture corresponds to that of the eukaryotic chloroplast, which has inner and outer envelope membranes and thylakoid membranes.

The major glycerolipids of cyanobacterial cells are monogalactosyl diacylglycerol (MGDG), digalactosyl diacylglycerol (DGDG), sulfoquinovosyl diacylglycerol (SQDG), and phosphatidylglycerol (PG) (2), and in eukaryotic chloroplasts these same glycerolipids predominate (3,4). MGDG accounts for about half of the total glycerolipids, and the other three glycerolipids contribute to the remaining half to different degrees, depending on the strain and specific growth conditions (5,6).

Cyanobacterial strains can be classified into four groups by reference to the unsaturation of fatty acids (Table 3.1). Group 1 is characterized by the presence of saturated and monounsaturated fatty acids excessively, whereas groups 2, 3 and 4 contain polyunsaturated fatty acids. The latter fatty acids are unusual in that the C_{18} and C_{16} fatty acids are esterified to the *sn*-1 and *sn*-2 positions of the glycerol moiety, respectively. Strains in group 1 [e.g., *Synechococcus* sp. PCC 7942 (*Anacystis nidulans* R2) and *Mastigocladus laminosus*] introduce a double bond only at the Δ9 position of fatty acids, either at the *sn*-1 or the *sn*-2 position (5,7). Strains in group 2 (e.g., *Synechococcus* sp. PCC 7002, *Anabaena variabilis*, *Plectonema boryanum*, and *Nostoc muscorum*) can introduce double bonds at the Δ9, Δ12 and Δ15 (ω-3) positions of C_{18} fatty acids at the *sn*-1 position, as well as at the Δ9 and Δ12 positions of C_{16} fatty acids at the *sn*-2 position (7,8). Strains in group 3 (e.g., *Synechocystis* sp. PCC 6714 and *Spirulina platensis*) can also introduce three double bonds, but these are found at the Δ6, Δ9 and Δ12 positions of C_{18} fatty acids at the *sn*-1 position (5). Strains in group 4 (e.g., *Synechocystis* sp. PCC 6803 and *Tolypothrix tenuis*) can introduce double bonds at the Δ6, Δ9, Δ12 and Δ15 (ω-3) positions of C_{18} fatty acids at the *sn*-1 position (5,9). However, desaturation at the *sn*-2 position in groups 1 and 2 and Δ6 desaturation at the *sn*-1 position in groups 3 and 4 are confined to MGDG, with both SQDG and PG being excluded. It is likely that desaturation does not

occur in fatty acids that are bound to DGDG; the various molecular species of DGDG are probably synthesized by galactosylation of the corresponding molecular species of MGDG (8). Among the members of the four groups of cyanobacteria, those in group 2 are the most similar to plant chloroplasts in terms of the desaturation of fatty acids.

Biosynthesis of γ-Linolenic Acid in *Spirulina*

The cyanobacterial strains in group 3 and group 4 synthesize γ-linolenic acid, while *Spirulina platensis*, which belongs to group 3, was found to contain the highest level of γ-linolenic acid among the strains examined (Table 3.1). We analyzed the fatty acid composition of individual lipid classes of *Spirulina platensis* (Table 3.2). The level of 16:0 ranged from 50 to 60% of the total fatty acids in each lipid class. γ-Linolenic acid was confined to MGDG and DGDG, whereas 18:2(9,12) was a main contributor of fatty acids to SQDG and DGDG.

The distribution of fatty acids at the *sn*-1 and *sn*-2 positions of the glycerol moieties of lipids (data not shown) indicated that the *sn*-2 position was exclusively esterified by 16:0, and that all of the C_{18} fatty acids and 16:1(9) were located at the *sn*-1 position. These findings suggest a pathway for the biosynthesis of fatty acids (or molecular species), as shown in Fig. 3.1. In MGDG, SQDG, and PG, the major precursors of all of the molecular species are *sn*-1-18:0/*sn*-2-16:0 species. In these precursors, the

TABLE 3.1 Major Fatty Acids of the Total Lipids from Various Strains of Cyanobacteria

Organism	Growth temp. (°C)	Fatty acid (mole %) 16:0	16:1 (9)	16:2 (9,12)	18:0	18:1 (9)	18:2 (9,12)	α18:3 (9,12,15)	γ18:3 (6,9,12)	18:4 (6,9,12,15)
Group 1										
Mastigocladus laminosus (F)	34	34	31	0	5	29	0	0	0	0
Synechococcus PCC7942 (U)	34	49	36	0	4	10*	0	0	0	0
Synechococcus PCC6301 (U)	38	48	38	0	4	7*	0	0	0	0
Synechococcus lividus (U)	38	42	36	0	1	20	0	0	0	0
Group 2										
Plectonema boryanum (F)	28	36	22	0	1	3*	10	29	0	0
Nostoc muscorum (F)	28	41	14	0	1	2	7	35	0	0
Anabaena variabilis (F)	22	29	22	3	t	7	15	24	0	0
Synechococcus PCC7002 (U)	22	35	19	0	t	10	25	10	0	0
Group 3										
Spirulina platensis (F)	32	53	3	0	1	1	13	0	29	0
Synechocystis PCC6714 (U)	34	59	2	0	t	9	16	0	12	0
Group 4										
Tolypothrix tenuis (F)	30	55	3	0	1	2	5	6	11	17
Synechocystis PCC6803 (U)	22	51	3	0	1	2	6	8	21	8

t: Trace (less than 0.5%). *Mixture of Δ9-octadecenoic acid (oleic acid) and Δ11-octadecenoic acid (*cis*-vaccenic acid). F and U in parentheses indicate filamentous and unicellular strains, respectively.

TABLE 3.2 Major Fatty Acids of the Various Lipid Classes in *Spirulina platensis* Grown at 34°C

Lipid class	Fatty acid (mole %) 16:0	16:1 (9)	18:1 (9)	18:2 (9,12)	γ18:3 (6,9,12)
MGDG (47%)	52	3	1	1	42
DGDG (16%)	51	5	2	3	38
SQDG (17%)	60	2	7	26	1
PG (20%)	55	1	5	35	1

first double bond is introduced at the Δ9 position of 18:0, at the *sn*-1 position, to yield 18:1(9), and the second double bond is introduced at the Δ12 position of 18:1(9), at the *sn*-1 position, to yield 18:2(9,12). In SQDG and PG, no further desaturation occurs. In MGDG, by contrast, a third double bond is introduced at the Δ6 position of 18:2(9,12) to yield 18:3(6,9,12), in other words, γ-linolenic acid.

It is likely that the molecular species of DGDG are synthesized by galactosylation of the corresponding molecular species of MGDG and that no desaturation takes place in DGDG itself (8). The scheme in Fig. 3.1 suggests the possible existence of three different types of desaturase, namely, Δ9 desaturase, Δ12 desaturase and Δ6

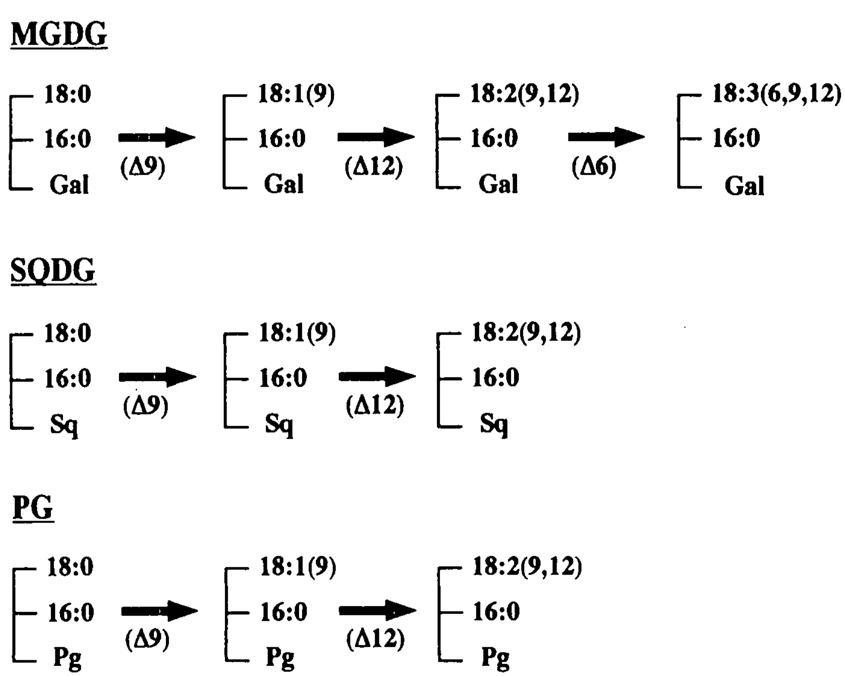


Fig. 3.1. Pathway for the desaturation of fatty acids in *Spirulina platensis*. Abbreviations: Gal, galactose; Sq, sulfoquinovose; Pg, glycerolphosphate

desaturase. This last desaturase is specific to MGDG, whereas the former two desaturases can use SQDG and PG as their substrates, in addition to MGDG.

There are three classes of desaturases. Acyl-CoA desaturases introduce double bonds into fatty acids bound to coenzyme A; these enzymes are bound to the endoplasmic reticulum in animal, yeast and fungal cells (10). Acyl-ACP desaturases introduce double bonds into fatty acids that are bound to ACP; they are present in the stroma of plant plastids (11). Acyl-lipid desaturases introduce double bonds into fatty acids that have been esterified to glycerolipids (12–14); they are bound to the endoplasmic reticulum, the chloroplast membrane in plant cells (12), and the thylakoid membranes in cyanobacterial cells (13). This last class of desaturase is the most efficient regulator of the extent of unsaturation of membrane lipids in response to changes in temperature.

The acyl-lipid desaturases can be further classified into two subgroups by reference to their electron donors. One subgroup, present in the endoplasmic reticulum of plant cells, uses cytochrome b_5 as the electron donor (15,16). The other, present in the chloroplasts of plant cells and in cyanobacterial cells, uses ferredoxin as the electron donor (13,17,18). A unique characteristic of the acyl-lipid desaturases is that they recognize, by an unknown mechanism, exactly those positions within various carbon chains at which double bonds are to be specifically introduced.

Cyanobacterial Desaturases

All known cyanobacterial desaturases are of the acyl-lipid and membrane-bound type (19). Because purification of these enzymes by conventional methods has proved difficult, we attempted the molecular cloning of the various desaturases. A mutant that was defective in the synthesis of 18:2(9,12), 18:3(6,9,12), 18:3(9,12,15), and 18:4(6,9,12,15) was initially isolated from *Synechocystis* sp. PCC 6803 after treatment of wild-type cells with ethyl methanesulfonate (9). The mutant, designated Fad12, was defective in desaturation at the Δ12 position of C_{18} fatty acids at the *sn*-1 position of the glycerol moiety in all lipid classes. The growth rate at 22°C of the mutant was much lower than that of the wild type, whereas mutant and wild-type cells grew at about the same rate at 34°C (20).

A gene (*desA*) for the Δ12 desaturase was isolated (21) by screening of the genomic DNA library of *Synechocystis* sp. PCC 6803 for the ability to complement the Fad12 mutation with respect to both growth at low temperature and desaturation at the Δ12 position of fatty acids after *in situ* transformation. The *desA* gene contains an open reading frame of 1053 bp that corresponds to 351 amino acid residues and encodes an acyl-lipid desaturase. The enzyme can introduce a second *cis*-double bond at the Δ12 position of fatty acids bound to membrane glycerolipids. Similar *desA* genes were isolated by heterologous hybridization from *Synechococcus* sp. PCC 7002, *Synechocystis* sp. PCC 6714 and *Anabaena variabilis* with a probe derived from the *desA* gene of *Synechocystis* sp. PCC 6803 (22). The amino acid sequence deduced from the nucleotide sequence of the *desA* gene of *Synechocystis* sp. PCC 6803 is similar to that of the gene from *Synechocystis* sp. PCC 6714. The extent of sequence similarity between the amino acid sequences from *Synechocystis*

sp. PCC 6803 and *Synechocystis* sp. PCC 6714 is 96%. However, the extent of conservation of the amino acid level between the sequences of polypeptides from *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 7002 and that between the sequences of polypeptides from *Synechocystis* sp. PCC 6803 and *A. variabilis* are 57 and 59%, respectively (22).

The *desC* gene for a Δ9 acyl-lipid desaturase was found in the 5′-upstream region of the *desA* gene on the chromosome of *A. variabilis* (23). The *desC* gene of *Synechocystis* sp. PCC 6803 was cloned by screening a genomic library with a probe derived from the *desC* gene of *A. variabilis*. The deduced amino-acid sequences of the Δ9 acyl-lipid desaturases of *Synechocystis* sp. PCC 6803 and *A. variabilis* are similar to those of the Δ9 acyl-CoA desaturases from rat (24), mouse (25,26), and yeast (27), with homology at the amino acid level of about 25% in each case.

We have cloned the *desB* gene for the ω-3 acyl-lipid desaturase from *Synechocystis* sp. PCC 6803 by screening a genomic library with a probe derived from the *desA* gene of this strain (28). Reddy et al. (29) cloned the *desD* gene for Δ6 desaturase by the "gain-of-function" method using *Anabaena* sp. PCC 7120, which does not contain a Δ6 desaturase. Currently genes for all of the desaturases have been cloned from *Synechocystis* sp. PCC 6803.

Molecular Cloning of Desaturases of *Spirulina*

The desA *Gene for Δ12 Desaturase*

From comparisons of the sequences of the *desA* genes from *A. variabilis*, *Synechocystis* sp. PCC 6803, *Synechocystis* sp. PCC 6714 and *Synechococcus* sp. PCC 7002 (22), we selected two conserved regions, namely, amino acid positions 116–121 and 233–237, counted from the amino terminus of the *desA* gene of *Synechocystis* sp. PCC 6803. Oligonucleotides corresponding to these regions, including a synthetic *Eco RI* restriction site at each 5′ end, were synthesized. Their sequences were as follows:

> 5-GGGAATTCTA(TC)CC(ACGT)TT(TC)CA(TC)AG(CT)TGG-3′ and
> 5′-GGGAATTCAC(AG)TT(AGT)AT(AG)TC(AG)TG(AG)CA-3′.

The oligonucleotides were used as forward and reverse primers, respectively, for amplification by the polymerase chain reaction (PCR) of a partial nucleotide sequence of the *desA* gene of *Spirulina platensis*, with genomic DNA as the template. The amplified products of about 400 bp were of the size predicted for the partial sequence of the *desA* gene of *Synechocystis* sp. PCC 6803. These products of PCR were subcloned into pBluescript(SK+) (Stratagene, La Jolla, CA) and nucleotide sequences were determined. A clone with a sequence homologous to that of the *desA* gene of *Synechocystis* sp. PCC 6803 was identified. A genomic DNA library of *S. platensis*, constructed in the phage vector *λDASHII* (Stratagene) was screened with the insert of this clone, and the *desA* gene for Δ12 desaturase was cloned (its sequence has been deposited to the EMBL Data Library with the accession number X86736). Figure 3.2 shows a comparison of the amino acid sequence deduced from the *desA* gene of *S. platensis* with those deduced from the *desA* genes of *A. variabilis*, *Synechocystis* sp. PCC 6803, *Synechocystis* sp. PCC 6714 and *Synechococcus* sp. PCC 7002. The extent of homology at the amino acid level between the *desA* gene of *S. platensis* and those of *A. variabilis*, *Synechocystis* sp. PCC 6803, *Synechocystis* sp.


```
Spirulina    MTLSIVKSEDSSSRPSAVPSDLPLEEDIINTLPSGVFVQDRYKAWMTV  48
Anabaena          MTTSTIKNQEIKNLSNPELRLKDILDTLPRSVYQQNRRKAWTQA  44
PCC 6803    MTATIPPLTPTVTPSNPDRPIADLKLQDIIKTLPKECFEKKASKAWASV  49
PCC 6714    MTATIPPLRPTETSSNPDRPIADLKLQDIIKTLPKECFEKKASKAWASV  49
PCC 7002        MTSVTVRPSATTLLEKHPNLRLRDILDTLPRSVYEINPLKAWSRV  45

Spirulina   IINVVMVGLGWLGIAIAPWFLLPVVWVFTGTALTGFFVIGHDCGHRSFSR  98
Anabaena    LLSVVMVGLCYWSLAIAPWFLLLPAWFFTGTTLTGFFVIGHDCGHRSFSR  94
PCC 6803    LITLGAIAVGYLGIIYLPWYCLPITWIWTGTALTGAFVVGHDCGHRSFAK  99
PCC 6714    LITLGAIALGYLGIIYLPWYCLPFTWIWTGTALTGAFVVGHDCGHRSFAK  99
PCC 7002    LLSVAAVVGCYALLAIAPWYLLLPVWFLTGTTLTGFFVIGHDCGHRSFSR  95
                                                    *   *
                                                    *   *

Spirulina   NVWVNDWVGHILFLPIIYPFHSWRIGHNQHHKYTNRMELDNAWQPWRKE-  147
Anabaena    RNWVNNLVGHLFMMPLIYPFHSWRIKHNHHHKYTNNLDEDNAWHPIRPEV  144
PCC 6803    KRWVNDLVGHIAFAPLIYPFHSWRLLHDHHHLHTNKIEVDNAWDPWSVE-  148
PCC 6714    KRWVNDLVGHIAFAPLIYPFHSWRLLHDHHHLHTNKIEVDNAWDPWSVE-  148
PCC 7002    KNWVNNLVGHLAFLPLIYPFHSWRILHNHHHRYTNMNDEDNAWAPFTPEL  145
                     *          *     *  **
                                      *  **

Spirulina   EYQNAGKFMQVTYDLFRGRAWWIGSILHWASIHFDWTKFEGKQRQQVKFS  197
Anabaena    YASWGKTRQSAFKLFMRQRLWWVASVGHQAVVHFDWRKFKVKQQADVRFS  194
PCC 6803    AFQASPAIVRLFYRAIRGPFWWTGSIFHWSLMHFKLSNFAQRDRNKVKLS  198
PCC 6714    AFQASPAIVRLFYRAIRGPFWWTGSIFHWGLMHFKLSNFAERDRNKVKLS  198
PCC 7002    YDDSPAFIKAVYRA-IRGKLWWLASVIHQLKLHFMWFAFEGKQREQVRFS  194
                                       *    *

Spirulina   SLLVIGAAAIAFPTMILTIGVWGF-VKFWVIPWLVFHFWMSTFTLLHHTI  246
Anabaena    FSLVVIAGAVAFPTMFATLGIWGFFVKFWFVPWLGYHFWMSTFTLVHHTY  244
PCC 6803    IAVVFLFAAIAFPALIITTGVWGF-VKFWLMPWLVYHFWMSTFTIVHHTI  247
PCC 6714    IAVVFLFAAVAFPALIITTGVWGF-VKFWLMPWLVYHFWMSTFTIVHHTI  247
PCC 7002    ALFVIIAGAIAPPVMFYGLGVWG-VVKFWLMPWLGYHFWMSTFTLVHHTV  243
                                                *         **

Spirulina   ADIPFREPEQWHEAESQLSGTVHCNYSRWGEFLCHDINVHIPHHVTTAIP  296
Anabaena    PDVPFEAENKWHEAMAQLFGTIHCDYPKWVEVLCHDINVHVPHHLSTGIP  294
PCC 6803    PEIRFRPAADWSAAEAQLNGTVHCDYPRWVEVLCHDINVHIPHHLSVAIP  297
PCC 6714    PEIRFRPAEDWSAAEAQLNGTVHCDYPRWVEVLCHDINVHIPHHLSVAIP  297
PCC 7002    PEIPFSYRDKWNEAIAQLSGTVHCDYPKWVEVLCHDINVHVPHHLHTGIP  293
                                  *           *    *  **
                                                   *  **

Spirulina   WYNLRTPTPVYRKIGGEYLYPECDFSWGLMKQVVDHAICMMRITIIS-QS  345
Anabaena    SYNLRKAYSSIQQNWGDYL-HELRFSWSLM-K-LITDECQLYQTDVNYQP  341
PCC 6803    SYNLRLAHGSLKENWGPFLY-ERTFNWQLMQQISGQ--CHLYDPEHGYRT  344
PCC 6714    SYNLRLAHASLKQNWGPFLY-ERTFNWGLMQQISGQ--CHLYDPDNGYRT  344
PCC 7002    SYNLRKALASIKQNWGEYL-YETKFHWELM-K-AITEQCHLYVAEHNYIS  340

Spirulina   LTT-KRV                                            351
Anabaena    FKDYYAGR                                           349
PCC 6803    FGSLKKV                                            351
PCC 6714    FSSLK                                              349
PCC 7002    FAQ-HQKR                                           347
```

Fig. 3.2. Alignment of the amino acid sequence of the Δ12 desaturase of *Spirulina platensis* with those of Δ12 desaturases of *Anabaena variabilis*, *Synechocystis* sp. PCC 6803, *Synechocystis* sp. PCC 6714, and *Synechococcus* sp. PCC 7002 (22). Histidine residues conserved in the Δ12 desaturases are marked by asterisks. The histidine residues conserved in the Δ6, Δ9, Δ12, and Δ15 (ω-3) desaturases (19) are marked by double asterisks.

PCC 6714 and *Synechococcus* sp. PCC 7002 was calculated to be 56.5, 52.3, 52.3 and 57.2%, respectively. Seventeen histidine residues are conserved in the five Δ12 desaturases. Among them, eight histidine residues (Fig. 3.2) are conserved in the four desaturases that act at the Δ6, Δ9, Δ12, and ω-3 positions (19). As demonstrated in other enzymes with non-heme iron as the catalytic center, such as rubrerythrin (30), isopenicillin N synthase (31), stearoyl-acyl carrier protein desaturase (32), and lipoxygenase (33,34), it seems very likely that these histidine residues provide ligands to the iron that acts as the catalytic center.

The desD *Gene for Δ6 Desaturase*

The open reading frame (ORF) of the *desD* gene of *Synechosystis* sp. PCC 6803 was amplified by PCR with 5'-ATGCTAACAGCGGAAAGAAT-3' and 5'-GATGCTTTGCCCATGGCCTC-3' as the forward and reverse primers, respectively. The amplified product was subcloned into the TA cloning site of pCRII (Invitrogen, San Diego, CA). The genomic DNA library of *Spirulina platensis* that had been constructed in *λDASHII* was screened with a probe derived from the amplified DNA product, and the *desD* gene

```
Spirulina   MTSTTSKVTFGKSIGFRKELNRRVNAYLEAENISPRDNPPMYLKTAIILA  50
PCC 6803     MLTAERIKFTQKRGFRRVLNQRVDAYFAEHGLTQRDNPSMYLKTLIIVL  49

Spirulina   WVVSAWTFVVFGPDVLWMKLLGCIVLGFGVSAVGFNISHDGNHGGYSKYQ 100
PCC 6803    WLFSAWAFVLFAPVIFPVRLLGCMVLAIALAAFSFNVGHDANHNAYSSNP  99
                                                      *   *
                                                      *   *

Spirulina   WVNYLSGLTHDAIGVSSYLWKFRHNVLHHTYTNILGHDVEIHGDELVRMS 150
PCC 6803    HINRVLGMTYDFVGLSSFLWRYRHNYLHHTYTNILGHDVEIHGDGAVRMS 149
                                    *  **        *    *
                                    *  **

Spirulina   PSMEYRWYHRYQHWFIWFVYPFIPYYWSIADVQTMLFKRQYHDHEIPSPT 200
PCC 6803    PEQEHVGIYRFQQFYIWGLYLFIPFYWFLYDVYLVLNKGKYHDHKIPPFQ 199
                                                      * *

Spirulina   WVDIATLLAFKAFGVAVFLIIPIAVGYSPLEAVIGASIVYMTHGLVACVV 250
PCC 6803    PLELASLLGIKLLWLGYVFGLPLALGFSIPEVLIGASVTYMTYGIVVCTI 249

Spirulina   FMLAHVIEPAEFLDPD-NL-HIDDEWAIAQVKTTVDFAPNNPIINWYVGG 298
PCC 6803    FMLAHVLESTEFLTPDGESGAIDDEWAICQIRTTANFATNNPFWNWFCGG 299
                *

Spirulina   LNYQTVHHLFPHICHIHYPKIAPILAEVCEEFGVNYAVHQTFFGALAANY 348
PCC 6803    LNHQVTHHLFPNICHIHYPQLENIIKDVCQEFGVEYKVYPTFKAAIASNY 349
                  **     * *
                  **

Spirulina   SWLKKMSINPETKAIEQLTV                               368
PCC 6803    RWLEAMG-----KAS                                    359
```

Fig. 3.3. Alignment of the amino acid sequence of the Δ6 desaturase of *Spirulina platensis* with that of *Synechocystis* sp. PCC 6803 (29). Conserved histidine residues are marked by asterisks. The histidine residues conserved in the Δ6, Δ9, Δ12, and Δ15 (ω-3) desaturases (19) are marked by double asterisks.

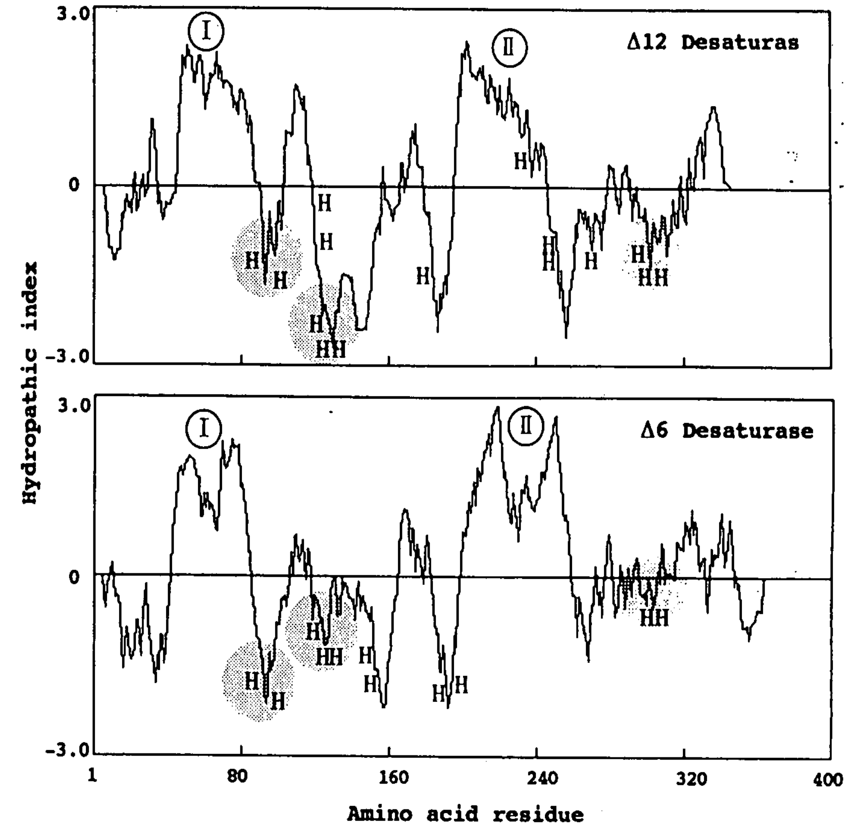


Fig. 3.4. Hydropathy plots and sites of histidine clusters in the Δ12 desaturase and Δ6 desaturase of *Spirulina platensis*. The three histidine (H) clusters conserved in the Δ6, Δ9, Δ12, and Δ15 (ω-3) desaturases (19) are shaded.

of *S. platensis* was cloned (its sequence has been deposited in the EMBL Data Library with the accession number X87094). Figure 3.3 shows the alignment of the deduced amino acid sequence of the Δ6 desaturase of *S. platensis* with that of *Synechocystis* sp. PCC 6803. The homology of the amino acid level between the two Δ6 desaturases is 52%, and 14 histidine residues are conserved in the two Δ6 desaturases. Seven of these histidine residues are conserved in all of the four desaturases that act at the Δ6, Δ9, Δ12, and ω-3 positions.

Figure 3.4 shows the predicted hydropathy profiles of the Δ12 and Δ6 desaturases of *S. platensis*. There are two hydrophobic domains, I and II, in both desaturases. It has been suggested that each domain spans the membrane twice, thus, each individual desaturase spans the membrane four times (19). Histidine residues are distributed throughout the sequences. However, the three histidine clusters conserved in all desaturases (19) are located at similar positions in both desaturases of *S. platensis*, and it seems likely that they are located on the cytoplasmic side of the membrane.

General Characteristics of *Spirulina platensis*

Spirulina platensis is a filamentous cyanobacterium, and its cells can be easily collected on nylon mesh. *S. platensis* grows under alkaline conditions with maximal growth at pH 10–11. These characteristics allow us to cultivate *S. platensis* in an open system and to collect the filamentous cells with a net, thereby freeing them of contaminating bacteria and fungi.

The history of *S. platensis* as a staple food for humans is of great interest (35). The cells contain a high proportion of protein (70% dry weight), several vitamins, and essential n-6 fatty acids. The strain that we used contains γ-linolenic acid, 18:3(6,9,12), as a major fatty acid. This compound has many pharmaceutical properties. It can relieve premenstrual syndrome (36) and is used as a treatment for a topic eczema. It also affects hyperlipidemia, which is frequently related to the development of arteriosclerosis and coronary heart disease, by lowering plasma levels of cholesterol and triglycerides.

Acknowledgments

The work summarized here was supported, in part, by Grants-in-Aid for Scientific Research on Priority Area (nos. 04273102 and 04273103) from the Ministry of Education, Science and Culture, Japan, to N.M.

References

1. Stanier, R.Y., and Cohen-Bazire, G. Phototrophic Prokaryotes: The Cyanobacteria (1977) *Annu. Rev. Microbiol. 31,* 225–274.
2. Omata, T., and Murata, N. Isolation and Characterization of the Cytoplasmic Membranes from the Blue-Green Alga (Cyanobacterium) *Anacystis nidulans* (1983) *Plant Cell Physiol. 24,* 1101–1112.
3. Block, M.A., Dorne, A.J., Joyard, J., and Douce, R. Preparation and Characterization of Membrane Fractions Enriched in Outer and Inner Envelope Membranes from Spinach Chloroplasts (1983) *J. Biol. Chem. 258,* 13281–13286.
4. Joyard, J., Block, M.A., and Douce, R. Molecular Aspects of Plastid Envelope Biochemistry (1991) *Eur. J. Biochem. 199,* 489–509.
5. Murata, N., Wada, H., and Gombos, Z. Modes of Fatty-Acid Desaturation in Cyanobacteria (1992) *Plant Cell Physiol. 33,* 933–941.
6. Murata, N., and Nishida, I. (1987) in *The Biochemistry of Plants,* Stumpf, P.K., Academic Press, Orlando, FL, vol. 9, pp. 315–347.
7. Sato, N., Murata, N., Miura, Y., and Ueta, N. Effect of Growth Temperature on Lipid and Fatty Acid Compositions in the Blue-Green Algae, *Anabaena variabilis* and *Anacystis nidulans* (1979) *Biochim. Biophys. Acta 572,* 19–28.
8. Sato, N., and Murata, N. Lipid Biosynthesis in the Blue-Green Alga, *Anabaena variabilis* II. Fatty Acids and Lipid Molecular Species (1982) *Biochim. Biophys. Acta 710,* 279–289.
9. Wada, H., and Murata, N. *Synechocystis* PCC 6803 Mutants Defective in Desaturation of Fatty Acids (1989) *Plant Cell Physiol. 30,* 971–978.

10. Holloway, P.W. (1983) in *The Enzymes,* Boyer, P. D., Academic Press, Orlando FL, vol. XVI, pp. 63–83.
11. McKeon, T.A., and Stumpf, P.K. Purification and Characterization of the Stearoyl-Acyl Carrier Protein Desaturase and the Acyl-Acyl Carrier Protein Thioesterase from Maturing Seeds of Safflower (1982) *J. Biol. Chem. 257,* 12141–12147.
12. Jaworski, J.G. (1987) in *The Biochemistry of Plants,* Stumpf, P.K., Academic Press, Orlando FL, vol. 9, pp. 159–174.
13. Wada, H., Schmidt, H., Heinz, E., and Murata, N. *In vitro* Ferredoxin-Dependent Desaturation of Fatty Acids in Cyanobacterial Thylakoid Membranes (1993) *J. Bacteriol. 175,* 544–547.
14. Sato, N., Seyama, Y., and Murata, N. Lipid-Linked Desaturation of Palmitic Acid in Monogalactosyl Diacylglycerol in the Blue-Green Alga (Cyanobacterium) *Anabaena variabilis* Studied *in vivo* (1986) *Plant Cell Physiol 27,* 819–835.
15. Kearns, E.V., Hugly, S., and Somerville, C.R. The Role of Cytochrome b_5 in Δ12 Desaturation of Oleic Acid by Microsomes of Safflower (*Carthamus tinctorius* L.) (1991) *Arch. Biochem. Biophys. 284,* 431–436.
16. Smith, M.A., Cross, A.R., Jones, O.T.G., Griffiths, W.T., Stymne, S., and Stobart, K. Electron-Transport Components of the 1-Acyl-2-Oleoyl-*sn*-Glycero-3-Phosphocholine Δ^{12}-Desaturase (Δ^{12}-Desaturase) in Microsomal Preparations from Developing Safflower (*Carthamus tinctorius* L.) Cotyledons (1990) *Biochem. J. 272,* 23–29.
17. Schmidt, H., and Heinz, E. Involvement of Ferredoxin in Desaturation of Lipid-Bound Oleate in Chloroplasts (1990) *Plant Physiol. 94,* 214–220.
18. Schmidt, H., and Heinz, E. Desaturation of Oleoyl Groups in Envelope Membranes from Spinach Chloroplasts (1990) *Proc. Natl. Acad. Sci. U.S.A. 87,* 9477–9480.
19. Murata, N., and Wada, H. Acyl-Lipid Desaturases and Their Importance in the Tolerance and Acclimation to Cold of Cyanobacteria (1995) *Biochem. J. 308,* 1–8.
20. Wada, H., Gombos, Z., Sakamoto, T. and Murata, N. Genetic Manipulation of Fatty Acids in Membrane Lipids in the Cyanobacterium *Synechocystis* PCC6803 (1992) *Plant Cell Physiol. 33,* 535–540.
21. Wada, H., Gombos, Z., and Murata, N. Enhancement of Chilling Tolerance of a Cyanobacterium by Genetic Manipulation of Fatty Acid Desaturation (1990) *Nature 347,* 200–203.
22. Sakamoto, T., Wada, H., Nishida, I., Ohmori, M., and Murata, N. Identification of Conserved Domains in the Δ12 Desaturases of Cyanobacteria (1994) *Plant Mol. Biol. 24,* 643–650.
23. Sakamoto, T., Wada, H., Nishida, I., Ohmori, M., and Murata, N. Δ9 Acyl-Lipid Desaturases of Cyanobacteria (1994) *J. Biol. Chem. 269,* 25576–25580.
24. Thiede, M.A., Ozols, J., and Strittmatter, P. Construction and Sequence of cDNA for Rat Liver Stearyl Coenzyme A Desaturase (1986) *J. Biol. Chem. 261,* 13230–13235.
25. Kaestner, K., Ntambi, J.M., Kelly, T.J., Jr., and Lane, M.D. Differentiation-Induced Gene Expression in 3T3-L1 Preadipocytes (1989) *J. Biol. Chem. 264,* 14755–14761.
26. Ntambi, J.M., Buhrow, S.A., Kaestner, K.H., Christy, R.J., Sibley, E., Kelly, T.J., Jr., and Lane, M.D. Differentiation-Induced Gene Expression in 3T3-L1 Preadipocytes (1988) *J. Biol. Chem. 263,* 17291–17300.
27. Stukey, J.E., McDonough, V.M., and Martin, C.E. The *OLE1* Gene of *Saccharomyces cerevisiae* Encodes the Δ9 Fatty Acid Desaturase and Can Be Functionally Replaced by the Rat Stearoyl-CoA Desaturase Gene (1990) *J. Biol. Chem. 265,* 20144–20149.
28. Sakamoto, T., Los, D.A., Higashi, S., Wada, H., Nishida, I., Ohmori, M., and Murata, N. Cloning of ω3 Desaturase from Cyanobacteria and Its Use in Altering the Degree of Membrane-Lipid Unsaturation (1994) *Plant Mol. Biol. 26,* 249–263.

29. Reddy, A.S., Nuccio, M.L., Gross, L.M., and Thomas, T.L. Isolation of a Δ^6-Desaturase Gene from the Cyanobacterium *Synechocystis* sp. Strain PCC6803 by Gain-of-Function Expression in *Anabaena* sp. Strain PCC7120 (1993) *Plant Mol. Biol. 27*, 283–300.
30. Donald, M.K., Jr., and Benet, C.P. Intrapeptide Sequence Homology in Rubrerythrin from *Desulfovibrio vulgaris*: Identification of Potential Ligands to the Diiron Site *Biochem. Biophys. Res. Commum. 181*, 337–341.
31. Li-June, M., and Lawrence, Q., Jr., NMR Studies of the Active Site of Isopenicillin N Synthase, A Non-Heme Iron (II) Enzyme (1991) *Biochemistry 30*, 11653–11659.
32. Brian, G.F., Shanklin, J., Somerville, C., and Münck, E. Stearoyl-Acyl Carrier Protein Δ9 Desaturase from *Ricinus communis* Is a Diiron-Oxo Protein (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 2486–2490.
33. Chen, X.S., and Funk, C.D. Structure-Function Properties of Human Platelet 12-Lipoxygenase: Chimeric Enzyme and *in vitro* Mutagenesis Studies (1993) *FASEB J. 7*, 694–701.
34. Suzuki, H., Kishimoto, K., Yoshimoto, T., Yamamoto, S., Kanai, F., Ebina, Y., Miyatake, A., and Tanabe, T. Site-Directed Mutagenesis Studies on the Iron-Binding Domain and the Determinant for the Substrate Oxygenation Site of Porcine Leukocyte Arachidonate 12-Lipoxygenase (1994) *Biochem. Biophys. Acta 1210*, 308–316.
35. Richmond, A. (1986) *CRC Handbook of Microalgal Mass Culture*, CRC Press Inc., Boca Raton, FL, pp. 212–230.
36. Horrobin, D.F. The Role of Essential Fatty Acids and Prostaglandins in the Premenstrual Syndrome (1993) *J. Reprod. Med. 28*, 465–468.

Chapter 4

Enzymatic Enrichment of γ-Linolenic Acid from Black Currant Seed Oil

H.-J. Wille and J. Wang

Nestlé Research Center,
Lausanne, Switzerland

Introduction

The biological importance of γ-linolenic acid (GLA, 6,9,12-octadecatrienoic acid) of the n-6 series is well known (1,2). GLA is a precursor in the synthesis of series 1 prostaglandins and is created in vivo by enzymatic desaturation of linoleic acid (9,12-octadecadienoic acid). Certain factors such as stress, poor nutrition and aging impair the effectiveness of the 6-desaturase enzyme responsible for this conversion. Dietary supplementation with GLA is able to by-pass this impediment in the metabolic process.

Natural sources of GLA contain variable amounts of this acid (Table 4.1) but this rarely exceeds 25% and is even lower for oils other than borage oil. Thus there has always been an interest in producing higher concentrates of GLA. Different fractionation techniques have been developed to enrich GLA from natural sources. These include urea fractionation of fatty acids (3–7), separation on Y-zeolite and lipase-catalyzed reactions, such as selective hydrolysis of GLA-containing triacylglycerols (8), and selective esterification of GLA-containing fatty acid mixtures (9) derived from borage or evening primrose oil.

This paper reports an investigation concerning the ability of enzymes which catalyze the esterification of fatty acids to discriminate between α- and γ-linolenic acid. Another

TABLE 4.1 Average Fatty Acid Composition (%) of Several Main GLA-Containing Seed Oils

Fatty Acid	Borage	Blackcurrant	Evening Primrose
C16:0	9–11	6– 8	5– 7
C18:0	2– 4	1– 2	1– 2
C18:1, Δ9	14–18	9–13	5–10
C18:2, Δ9	35–40	44–51	73–78
C18:3, Δ6 (γ)	21–25	15–20	7–10
C18:3, Δ9 (α)	—	12–14	—
C18:4, Δ6	—	2– 4	—

Biochem. J. (1998) 330, 611–614 (Printed in Great Britain)

5

RESEARCH COMMUNICATION

Identification of a *Caenorhabditis elegans* Δ^6-fatty-acid-desaturase by heterologous expression in *Saccharomyces cerevisiae*

Johnathan A. NAPIER[1], Sandra J. HEY, Dominic J. LACEY and Peter R. SHEWRY
IACR-Long Ashton Research Station, Department of Agricultural Sciences, University of Bristol, Long Ashton, Bristol BS18 9AF, U.K.

We identified a cDNA expressed sequence tag from an animal (the nematode worm *Caenorhabditis elegans*) that showed weak similarity to a higher-plant microsomal Δ^6-desaturase. A full-length cDNA clone was isolated and expressed in the yeast *Saccharomyces cerevisiae*. This demonstrated that the protein encoded by the *C. elegans* cDNA was that of a fatty acid Δ^6-desaturase, as determined by the accumulation of γ-linolenic acid. The *C. elegans* Δ^6-desaturase contained an N-terminal cytochrome b_5 domain, indicating that it had a similar structure to that of the higher-plant Δ^6-desaturase. The *C. elegans* Δ^6-desaturase mapped to cosmid W08D2, a region of chromosome III. This is the first example of a Δ^6-desaturase isolated from an animal and also the first example of an animal desaturase containing a cytochrome b_5 domain.

INTRODUCTION

Over the last few years, a number of microsomal and soluble fatty acid desaturases have been isolated from higher plants, most notably *Arabidopsis thaliana* (thale cress). This has been achieved by a combined genetic and biochemical approach to the generation and complementation of mutant *Arabidopsis* lines defective in fatty acid desaturation or elongation [1]. The importance of this approach has been clearly validated by the isolation and characterization of genes encoding microsomal desaturases such the Δ^{12} [2] and Δ^{15} [3] (encoded by the *FAD2* and *FAD3* genes respectively) enzymes, which had previously proved intractable to classical purification techniques on account of their hydrophobicity. The isolation of these and related genes, such as the Δ^{12}-hydroxylase from *Ricinus communis* (castor bean) [4], has allowed the identification of a number of conserved motifs in plant microsomal desaturases, most notably the so called 'histidine boxes' [5]. These short motifs appear to be required for enzyme function and also allow the proteins containing these motifs to be classified as di-iron-centre-containing enzymes [6].

Recently we isolated a cDNA clone from borage (*Borago officinalis*), using highly degenerate PCR against these histidine motifs, which was shown by heterologous expression in transgenic tobacco (*Nicotiana tabacum*) to encode a microsomal Δ^6-desaturase [7]. Desaturation at the Δ^6 position is an unusual modification in higher plants, occurring only in a small number of species such as borage, evening primrose (*Oenothera* spp.) and redcurrant (*Ribes* spp.), which accumulate the Δ^6-unsaturated fatty acids γ-linolenic acid (GLA) and octadecatetraenoic acid in the seeds and/or leaves. GLA is a high-value plant fatty acid and is widely used in the treatment of a number of medical conditions, including eczema and mastalgia. It has been postulated that the application of GLA replaces the loss of endogenous Δ^6-unsaturated fatty acids [7]. The sequence of the borage microsomal Δ^6-desaturase differed from previously characterized plant microsomal desaturases/hydroxylases in that it contained an N-terminal extension which showed sequence similarity to cytochrome b_5, and also in that the third (most C-terminal) histidine box varied from the consensus [6] H-X-X-H-H, with a glutamine residue replacing the first histidine one.

Although Δ^6-fatty-acid desaturation is an unusual modification in higher plants, it is a common reaction in animals. The essential fatty acid linoleic acid ($C_{18:2,\Delta9,12}$) is desaturated to GLA by a Δ^6-desaturase as the first step on the biosynthetic pathway of the eicosanoids, which includes prostaglandins and leukotrienes. This results in the rapid metabolism of GLA [to dihomo-GLA ($C_{20:3,\Delta8,11,14}$) and arachidonic acid ($C_{20:4,\Delta5,8,11,14}$)], so accumulation of this fatty acid is not usually observed. For example, in the model animal system, the nematode *Caenorhabditis elegans*, polyunsaturated fatty acids which have been Δ^6-desaturated (in the form of arachidonic and eicosapentanoic acids) make up over 20% of the fatty acids of the total lipids, but no GLA is observed [8]. This is presumably due to its rapid elongation to C_{20} fatty acid derivatives.

We wished to determine whether the Δ^6-desaturase isolated from borage was representative of Δ^6-desaturases as a whole. Since most higher plants do not contain this enzyme [7], we decided to take advantage of the large amount of animal sequences available on public databases. To this end we identified a putative *C. elegans* Δ^6-desaturase expressed sequence tag (EST) and verified its function by expressing the corresponding cDNA in yeast. When the nematode coding sequence was expressed in yeast supplemented by the addition of linoleic acid, GLA was produced. This was confirmed by GC–MS, identifying the coding sequence similar to the *C. elegans* predicted open reading frame (ORF) W08D2.4 as a Δ^6-desaturase.

MATERIALS AND METHODS

The National Center for Biotechnology Information (NCBI) EST sequence database was searched for polypeptide sequences which were related to the higher-plant Δ^6-fatty-acid desaturase [7] and contained the variant histidine box Q-X-X-H-H. Putative positive *C. elegans* ESTs were further characterized by searching the *C. elegans* EST project database (http://www.ddbj.nig.ac.jp/

Abbreviations used: EST, expressed sequence tag; GLA, γ-linolenic acid; NCBI, National Center for Biotechnology Information; ORF, open reading frame.

[1] To whom correspondence should be sent.

htmls/c-elegans/html/ce-index.html) in order to identify related cosmid clones.

A partial cDNA clone identified by these searches was obtained from the *C. elegans* EST project (kindly supplied by Professor Y. Kohara, National Institute of Genetics, Mishima, Japan), and this was used to screen a *C. elegans* cDNA library (mixed stage; also supplied by Professor Y. Kohara) constructed in λZAPII. A number of positives were identified and further purified, and full-length clones were confirmed by sequencing to encode a transcript likely to have been transcribed from the gene designated W08D2.4, on cosmid W08D2, as determined by database searching of the genes sequenced by the *C. elegans* genome project.

The coding sequence of W08D2.4 was introduced into the yeast expression vector pYES2 by PCR. Oligonucleotides with 5′ overhangs were used to introduce *Kpn*I and *Sac*I sites at the 5′ and 3′ ends respectively. The fidelity of the construct was checked by *in vitro* transcription and translation using the *TnT* system (Promega).

The resulting plasmid was introduced into yeast (*Saccharomyces cerevisiae*) by the lithium acetate method [9], and expression of the transgene was induced by addition of galactose. The yeast was supplemented by the addition of 0.2 mM linoleate in the presence of 1 % tergitol, following the method of [10].

Yeast total fatty acids were analysed by GC of methyl esters, exactly as described previously [7]. Confirmation of the presence of GLA was carried out by GC–MS using a Kratos MS80RFA instrument operating at an ionization voltage of 70 eV, with a scan range of 500–40 Da. The mass spectrum of the novel peak resolved by GC was compared with that of an authentic GLA standard (Sigma).

RESULTS

The sequence of the borage Δ^6-desaturase was used to search databases for related sequences in species which, although they do not accumulate GLA, might be expected to perform Δ^6-desaturation. The simplest organism which fulfilled this criterion was the free-living nematode *C. elegans*. This small animal has been subject to both random cDNA (EST) sequencing programs and large-scale genome sequencing. Our searches of EST databases identified a high-scoring nematode EST, namely yk436b12. This partial sequence of 448 bases was used to search for related cosmid clones sequenced by the Japan *C. elegans* genome project, using the DNA database of the Japan *C. elegans* EST project server. This indicated that the clone yk436b12 showed sequence similarity to part of a gene present on cosmid W08D2 (GenBank accession number Z70271), which forms part of chromosome III [11]. Bases 21–2957 of cosmid W08D2 are predicted by the protein prediction program Genefinder [11] to encode an ORF of 473 residues which is interrupted by five introns. Examination of this predicted protein sequence (designated W08D2.4 by the Sanger Centre Nematode Sequencing Project, Hinxton, Saffron Walden, Essex, U.K.) revealed that it had a number of characteristics reminiscent of a microsomal fatty acid desaturase, including three histidine boxes. However, the predicated protein sequence indicated the presence of an N-terminal domain similar to that of cytochrome b_5, containing the diagnostic H-P-G-G motif found in cytochrome b_5 proteins [12]. Since the Δ^6-desaturase isolated by us from borage [7] also contained an N-terminal b_5 domain, this indicated that W08D2.4 may encode a Δ^6-desaturase. Closer examination of the sequence revealed the presence of the variant third histidine box, with a H → Q substitution (again as observed in the borage Δ^6-desaturase). However, the similarity between W08D2.4 and the borage Δ^6-desaturase is low (51.7 %), as is the value of 31.0 % for identity. Since W08D2.4 was encoded by a gene containing many introns, it was necessary to isolate a full-length cDNA to verify the sequence predicated by the Genefinder program [11] and also to allow the expression of the ORF to define the encoded function.

A cDNA library and EST yk436b12 were generously provided by Professor Y. Kohara, and a number of positive plaques were identified by screening with the EST insert. These were further purified to homogeneity, excised, and the largest inserts (~

```
Boofd6  MAAQIKKYIT SDELKNHDKP GDLWISIQGK AYDVS.DWVK DHPGGSFPLK SLAGQEVTDA  59
Ceeld6  .......... ...MVVDKNA SGLRMKVDGK WLYLSEELVK KHPGGAV.IE QYRNSDATHI  46

Boofd6  FVAFHPAS.. TWKNLDKF.. ...FTGYYLK DY........ ...SVSEVS. KDYRKLVFEF 100
Ceeld6  FHAFHEGSSQ AYKQLDLLKK HGEHDEFLEK QLEKRLDKVD INVSAYDVSV AQEKKMVESF 106

Boofd6  SKMGLYDKKG HIMFA..TLC FIAMLFAMSV YGVLFCEGVL VHLF.SGCLM GFLWIQSGWI 157
Ceeld6  EKLRQKLHDD GLMKANETYF LFKAISTLSI MAFAFYLQYL GWYITSACLL ALAWQQFGWL 166

Boofd6  GHDAGHYMVV SDSRLNKFMG IFAANCLSGI SIGWWKWNHN AHHIACNSLE YDPDLQYIPF 217
Ceeld6  THEFCHQQPT KNRPLNDTIS LFFGNFLQGF SRDWWKDKHN THHAATNVID HDGDIDLAP. 225

Boofd6  LVVSSKFFGS LTSHFYEKRL TFDSLSRFFV SYQHWTFYPI MCAARLNMYV QSLIMLLTKR 277
Ceeld6  ......LFAF IPGDLCKYKA SFEKAILKIV PYQHLYFTAM LPMLRFSWTG QSVQWVFKEN 279

Boofd6  NVSYRAHE.. .......LLG CLVFSIWYPL LVSCLPNWGE RIMFVIASLS VTGMQQVQ.F 327
Ceeld6  QMEYKVYQRN AFWEQATIVG HWAW.VFYQL FL..LPTWPL RVAYFIISQM GGGLLIAHVV 336

Boofd6  SLNHFSSSVY VGKPKG.NNW FEKQTDGTLD ISCPPWMDWF HGGLQFQIEH HLFPKMPRCN 386
Ceeld6  TFNHNSVDKY PANSRILNNF AALQILTTRN MTPSPFIDWL WGGLNYQIEH HLFPTMPRCN 396

Boofd6  LRKISPYVIE LCKKHNLPYN YASFSKANEM TLRTLRNTA. .LQARDITKP LPKNLVWEAL 444
Ceeld6  LNACVKYVKE WCKENNLPYL VDDYFDGYAM NLQQLKNMAE HIQAKAA*.. .......... 443

Boofd6  HTHG*448
Ceeld6  .....443
```

Figure 1 A comparison of the deduced amino acid sequences of the borage (*B. officinalis*) Δ^6-desaturase [7] and the *C. elegans* cDNA CeD6.1

Abbreviations: Ceeld6, CeD6.1; Boofd6, borage Δ^6-desaturase.

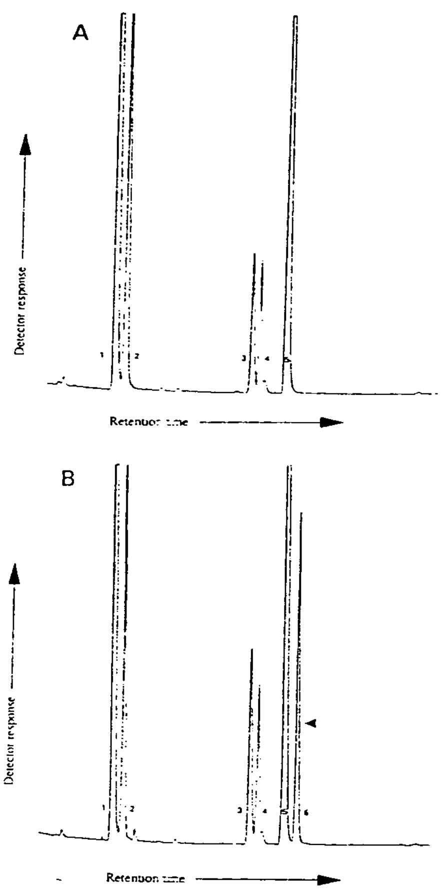


Figure 2 **Identification of GLA in transgenic yeast by GC**

[illegible] esters of total lipids of *S. cerevisiae* grown under inducing conditions (linoleate [illegible]) were analysed by GC, using flame-ionization detection. (**A**) is yeast transformed [illegible] empty) vector pYES2 and (**B**) is transformed with pYCeD6.1. The common peaks [illegible] as $C_{16:0}$ (peak 1), $C_{16:1}$ (peak 2), $C_{18:0}$ (peak 3), $C_{18:1}$ (peak 4) and $C_{18:2}$ (peak [illegible] exogenously). The additional peak (peak 6 in **B**), which corresponds to the retention time of GLA, is indicated by the arrowhead.

1450 bp) from the resulting rescued phagemids were sequenced. This confirmed that the cDNAs isolated by us did indeed show [illegible] to W08D2.4, with the 5′ and 3′ ends of the cDNA [illegible] equivalent to bases 9 and 3079 of the sequence of cosmid W08D2. Since the ATG initiating coding predicted by the Genefinder program to be the start of gene product W08D2.4 was indeed the first methionine residue in the cDNA clone, we reasoned that we had isolated a *bona fide* full-length cDNA. One representative cDNA clone (termed Cede.1; 1463 bp in length) was sequenced on both strands (Genbank ID: AF031477); the deduced amino acid sequence is identical with that predicted for W08D2.4 over the majority of the protein. However, DNA [illegible] encoding residues 38–67 (Y-S-I...L-Y-F) predicted for W[illegible] are not present in the cDNA clone. This means that the deduced amino acid sequence of pCeD6.1 is in fact 443 amino acids long, as opposed to that predicted for W08D2.4, which is 473 residues in length. The only other difference between the two amino acid sequences is an M → V substitution at residue 401, resulting from a A → G base change (base 1211). The deduced amino acid sequence of CeD6.1 is shown in Figure 1, compared with the previously characterized borage Δ^6-desaturase [7]. Note the presence in the *C. elegans* sequence of the H-P-G-G cytochrome b_5 motif in the N-terminus (encoded by bases 96–108) and the H → Q substitution in the third histidine box (encoded by bases 1157–1172).

Clone pCeD6.1 was then used as a template for PCR amplification of the entire predicated coding sequence (443 amino acid residues in length) and cloned into the yeast expression vector pYES2 (Invitrogen) to yield pYCeD6. The fidelity of this PCR-generated sequence was checked by *in vitro* transcription/translation of the plasmid, using the T_7 RNA polymerase promoter present in pYES2. Using the Promega *TnT*-coupled transcription/translation system, translation products were generated and analysed by SDS/PAGE and autoradiography, following the supplier's instructions. This revealed (results not shown) that the plasmid pYCeD6 generated a product of molecular mass 55 kDa, whereas the control (pYES2) failed to yield any protein products, indicating that the construct was correct.

Transformation and selection of yeast able to grow on uracil-deficient medium revealed yeast colonies carrying the recombinant plasmid pYCeD6 by virtue of the URA3-selectable marker carried by pYES2. Expression of pYCeD6 was obtained by inducing the GAL promoter which is present in pYES2. This was carried out after the cells had been grown up overnight with raffinose as a carbon source, and the medium supplemented by the addition of linoleate ($C_{18:2,\Delta9,12}$) in the presence of low concentrations of detergent. This latter addition was required since the normal substrate for Δ^6-desaturation is $C_{18:2}$ fatty acid, which does not normally occur in *S. cerevisiae*. The cultures were then allowed to continue to grow after induction, with aliquots being removed for analysis by GC. When methyl esters of total fatty acids isolated from yeast carrying the plasmid pYCeD6, grown in the presence of galactose and linoleate, were analysed by GC, an additional peak was observed (Figure 2). This had the same retention time as an authentic GLA standard, indicating that the transgenic yeast was capable of desaturating linoleic acid at the Δ^6 position. No such peaks were observed in any of the control samples (transformation with pYES2). The identity of this extra peak was confirmed by GC–MS, which positively identified the compound as GLA (Figure 3). This confirms that Cede.1 encodes a *C. elegans* Δ^6-desaturase, and that this cDNA is likely to be transcribed from the gene predicted to encode ORF W08D2.4, though the deduced amino acid sequence of Cede.1 is 30 residues smaller than that of W08D2.4.

DISCUSSION

Organisms such as *C. elegans* perform Δ^6-desaturation, but unlike plants such as borage or evening primrose, they do not accumulate Δ^6 unsaturated fatty acids such as GLA. We provide evidence that a *C. elegans* cDNA (Cede.1) encodes a Δ^6-desaturase, and that this sequence is similar to the predicted ORF W08D2.4, except for a 30-residue insertion present in the N-terminal region of the latter protein. Whether the deduced amino acid sequence predicted for Cede.1 represents a splicing variant of W08D2.4, or is a result of a misprediction of the intron/exon junctions by the Genefinder program is unclear. However, it is clear that Cede.1 encodes a Δ^6-desaturase. The ORF encoded by this *C. elegans* sequence appears to be related to the higher-plant Δ^6-fatty-acid desaturase previously isolated by us [7], in that they both contain N-terminal domains which show similarity to cytochrome b_5. In contrast, other microsomal fatty acid desaturases from plants do not contain this domain and use free cytochrome b_5 as an electron donor [1,13,14]. Similarly, the domain is absent from the only fatty acid

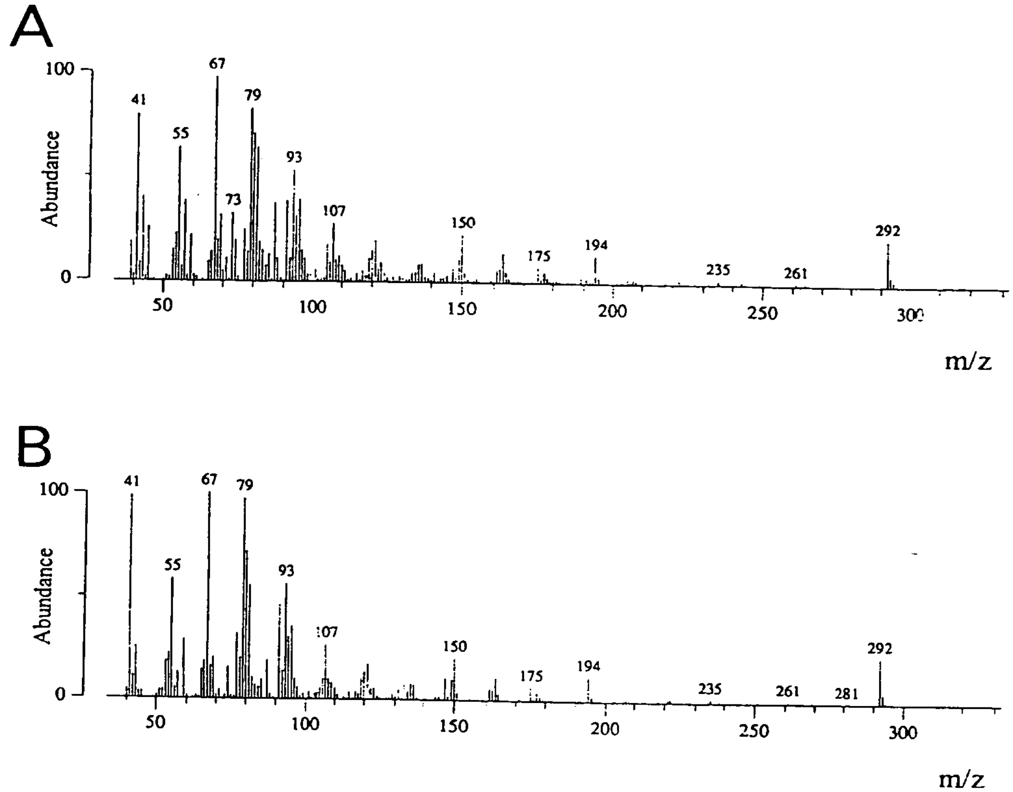


Figure 3 GC–MS analysis of the novel peak identified in yeast carrying pYCeD6.1

The sample was analysed for mass spectra as described previously [7], and the data were used to search a library of profiles. The sample was identified as GLA. A comparison of the mass spectra of the novel peak (**A**) and an authentic GLA standard (**B**) is shown. Visual- and computer-based inspection indicates that the two spectra are identical.

desaturases isolated from animals, a desaturase from *C. elegans* which recognizes a range of C_{18} and $C_{20,\omega-6}$ substrates [15] and a putative fatty acid desaturase from man (*Homo sapiens*) [16]. These animal sequences also differ from the borage and *C. elegans* Δ^6-desaturases in lacking the variant histidine box.

The reason why the Δ^6-desaturases have a fused cytochrome b_5 domain is not known [17]; the only other examples of desaturases with this extension are fungal microsomal (OLE1) Δ^9-desaturases [10] in which the domain is fused to the C-terminus rather than the N-terminus of the protein. However, the borage Δ^6-desaturase differs from all the other characterized plant microsomal desaturases in carrying out 'front-end' desaturation, which is the introduction of a double bond between C-3 and C-7 of an already unsaturated fatty acid [18]. This means the enzyme desaturates at positions between the carboxy group and pre-existing double bonds, whereas other plant enzymes desaturate sequentially towards the methyl group. It will be of interest to determine whether this feature is shared by other 'front-end' desaturases of plant and animal origin. It is also clear that identification of heterologous fatty acid desaturases will be facilitated by the yeast expression system described in the present study.

We are very grateful to Professor Yuji Kohara for supplying nematode ESTs and cDNA libraries. We thank Mervyn Lewis for carrying out the GC–MS analysis. IACR-Long Ashton Research Station receives grant-aided support from the Biotechnology and Biological Sciences Research Council (U.K.).

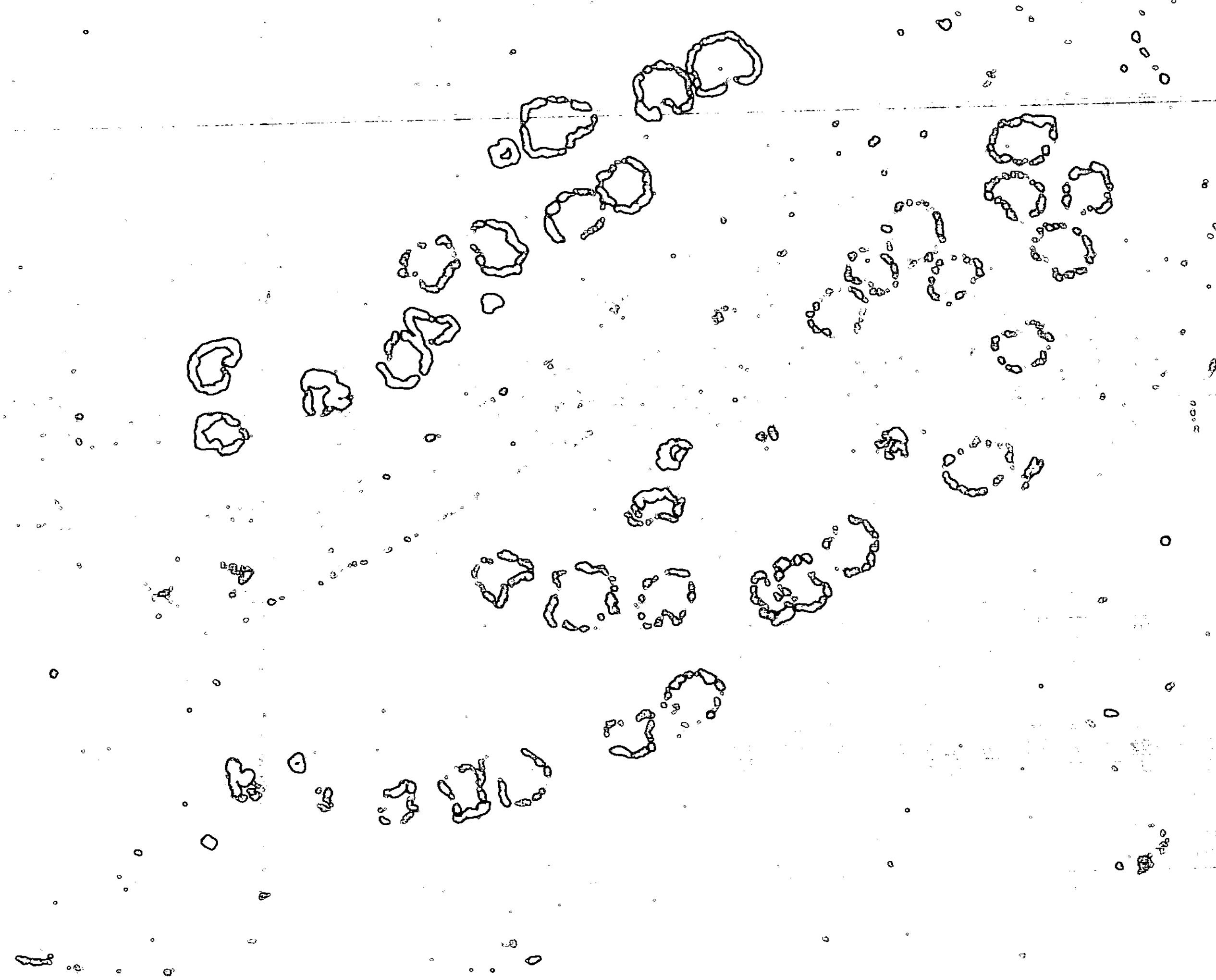
Received 5 December 1997/13 January 1998; accepted 13 January 1998

REFERENCES

1 Somerville, C. and Browse, J. (1996) Trends Cell Biol. **6**, 148–1153
2 Okuley, J., Lightner, J., Feldmann, K., Yadav, N., Lark, E. and Browse, J. (1994) Plant Cell **6**, 147–158
3 Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H. and Somerville, C. R. (1992) Science **258**, 1353–1355
4 van de Loo, F. N., Broun, P., Turner, S. and Somerville, C. R. (1995) Proc. Natl. Acad. Sci. U.S.A. **92**, 6743–6747
5 Shanklin, J., Whittle, E. and Fox, B. G. (1994) Biochemistry **33**, 12787–12794
6 Shanklin, J., Achim, C., Schmidt, H., Fox, B. G. and Munck, E. (1997) Proc. Natl. Acad. Sci. U.S.A. **94**, 2981–1986
7 Sayanova, O., Smith, M. A., Lapinskas, P., Stobart, A. K., Dobson, G., Christie, W. W., Shewry, P. R. and Napier, J. A. (1997) Proc. Natl. Acad. Sci. U.S.A. **94**, 4211–4216
8 Tanaka, T., Ikita, K., Ashida, T., Motoyama, Y., Yamaguchi, Y. and Satouchi, K. (1996) Lipids **31**, 1173–1178
9 Elble, R. (1992) Biotechniques **13**, 18–20
10 Mitchell, A. G. and Martin, C. E. (1995) J. Biol. Chem. **270**, 29766–29772
11 Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Burton, J., Connell, M., Bonfield, J., Copsey, T., Cooper, J. et al. (1994) Nature (London) **368**, 32–38
12 Lederer, F. (1994) Biochimie **76**, 674–692
13 Smith, M. A., Cross, A. R., Jones, O. T. G., Griffiths, W. T., Stymne, S. and Stobart, A. K. (1990) Biochem. J. **272**, 23–29
14 Smith, M. A., Jonsson, L., Stymne, S. and Stobart, A. K. (1992) Biochem. J. **287**, 141–144
15 Spychalla, J. P., Kinney, A. J. and Browse, J. (1997) Proc. Natl. Acad. Sci. U.S.A. **94**, 1142–1147
16 Cadena, D. L., Kurten, R. C. and Gill, G. N. (1997) Biochemistry **36**, 6960–6967
17 Napier, J. A., Sayanova, O., Stobart, A. K. and Shewry, P. R. (1997) Biochem. J. **328**, 717–718
18 Aitzetmüller, K. and Tseegsuren, N. (1994) J. Plant Physiol. **143**, 538–543

plant JOURNAL

4



Blackwell Science

lume 15
umber 1
uly 1998

The Plant Journal

Typeset and printed by Information Press Ltd, Oxford, UK

The Plant Journal (1998) 15(1), 39–48

Identification of a novel Δ6-acyl-group desaturase by targeted gene disruption in *Physcomitrella patens*

Thomas Girke[1], Hermann Schmidt[2], Ulrich Zähringer[3], Ralf Reski[4] and Ernst Heinz[1,*]

[1]*Universität Hamburg, Institut für Allgemeine Botanik, Ohnhorststr. 18, D-22609 Hamburg, Germany*

[2]*Institut f. Züchtung Landwirtschaftlicher Kulturpflanzen, Institutplatz 1, D-18190 Groß Lüsewitz, Germany*

[3]*LG Immunchemie, Parkallee 1–40, D-23845 Borstel, Germany, and*

[4]*Albert-Ludwigs-Universität Freiburg, Institut für Biologie II, Schänzlestr. 1, D-79104 Freiburg, Germany*

Summary

The moss *Physcomitrella patens* contains high levels of arachidonic acid. For its synthesis from linoleic acid by desaturation and elongation, novel Δ5- and Δ6- desaturases are required. To isolate one of these, PCR-based cloning was used, and resulted in the isolation of a full-length cDNA coding for a putatively new desaturase. The deduced amino acid sequence has three domains: a N-terminal segment of about 100 amino acids, with no similarity to any sequence in the data banks, followed by a cytochrome b_5-related region and a C-terminal sequence with low similarity (27% identity) to acyl-lipid desaturases. To elucidate the function of this protein, we disrupted its gene by transforming *P. patens* with the corresponding linear genomic sequence, into which a positive selection marker had been inserted. The molecular analysis of five transformed lines showed that the selection cartridge had been inserted into the corresponding genomic locus of all five lines. The gene disruption resulted in a dramatic alteration of the fatty acid pattern in the knockout plants. The large increase in linoleic acid and the concomitant disappearance of γ-linolenic and arachidonic acid in all knockout lines suggested that the new cDNA coded for a Δ6-desaturase. This was confirmed by expression of the cDNA in yeast and analysis of the resultant fatty acids by GC–MS. Only the transformed yeast cells were able to introduce a further double bond into the Δ6-position of unsaturated fatty acids. To our knowledge, this is the first report of a successful gene disruption in a multicellular plant resulting in a specific biochemical phenotype.

Intr duction

Compared to higher plants, many members of moss, algae and fern families produce a wider variety of polyunsaturated fatty acids (PUFA; Dembitsky, 1993; Jamieson and Reid, 1975; Zhukova and Aizdaicher, 1995), and PUFA such as arachidonic acid (AA) and eicosapentaenoic acid (EPA) are produced only by lower plants. The function of these long-chain PUFA in the membranes of lower plants is still unclear, whereas in humans, they play a key role in eicosanoid metabolism (Samuelsson, 1983).

The biosynthesis of AA and EPA generally starts with linoleic acid (18:2), which is channelled into a widely branching network of desaturation and elongation steps (Arao and Yamada, 1994; Cohen *et al.*, 1995; Shiran *et al.*, 1996). Key enzymes in this network are Δ5- and Δ6-desaturases, which introduce the new double bond between the first double bond and the carboxyl terminus of the fatty acid, known as carboxyl-directed desaturation. This mode differs from the methyl-directed desaturation, which works towards the methyl end of the unsaturated fatty acid. Desaturases of both types belong to the membrane-bound desaturases, which operate in microsomes or in plastids (Heinz, 1993). All desaturases, including acyl-ACP, (Ohlrogge *et al.*, 1993), acyl-CoA (Enoch *et al.*, 1976) and acyl-lipid desaturases, are believed to catalyse an O_2-dependent reaction, in which either cytochrome b_5 serves as electron donor for the microsomal or ferredoxin for the plastidial desaturases (Kearns *et al.*, 1991; Schmidt and Heinz, 1990; Smith *et al.*, 1990).

In the last few years, extensive sequence information from various desaturases in the methyl-directed group has been accumulated, but only a few from the carboxyl-directed group (Reddy *et al.*, 1993; Sayanova *et al.*, 1997) have been cloned so far. A good source to clone new desaturases is the moss *Physcomitrella patens*. Lipids of *P. patens* contain high proportions of AA (up to 30% of total fatty acids) indicating strong expression of Δ5- and Δ6-desaturases (Grimsley *et al.*, 1981). This moss can be propagated vegetatively in the haploid state (Ashton and Cove, 1977), which simplifies the phenotypic analysis after mutation or transformation (Schaefer *et al.*, 1991). Genes of this organism can be specifically inactivated by gene targeting, as shown by Schaefer and Zrÿd (1997), who demonstrated that integration of homologous DNA into the genome of *P. patens* takes place by homologous recombination with a relative efficiency of more than 90% among transgenic plants.

In the present communication, we describe the isolation of a new cDNA and its corresponding genomic sequence from *P. patens*, using a PCR-based screening. The encoded protein shared less than 27% sequence identity with known desaturases and represents a fusion between a C-terminal

Received 2 January 1998; revised 30 March 1998; accepted 8 April 1998.
*For correspondence (fax + 4 08228 2254; e-mail eheinz@botanik.uni-hamburg.de).

desaturase with a cytochrome b_5-related part and a N-terminal extension. Its function and importance for the biosynthesis of AA (20:4) was identified by disrupting the corresponding gene in *P. patens*. The biochemical phenotype of the null mutant and its subsequent complementation by feeding γ-linolenic acid ($18{:}3^{\Delta 6,9,12}$) demonstrated that the disrupted gene codes for a Δ6-desaturase, which plays a key role in the synthesis of 20:4.

Results

PCR-based cloning

For PCR experiments, different sets of degenerate primers, deduced from the three conserved histidine boxes of acyl-lipid desaturases, were synthesized (Avelange-Macherel *et al.*, 1995; Shanklin *et al.*, 1994). The template used was single-stranded cDNA from *P. patens*, which was reverse-transcribed from mRNA of 12-day-old protonema cultures. Bands of the expected length were cloned and sequenced. Data bank searches and alignments with these new sequences indicated similarities to acyl-lipid desaturases for seven cDNA fragments. Six of them were classified as putative members of the well-known Δ12- and Δ15-desaturases based on high identities of over 60%. In contrast to this, one sequence of 550 bp showed less than 27% identity to known desaturases. Since *Physcomitrella* was expected to express Δ5- and Δ6-desaturases, it was postulated that this sequence might be derived from one of those desaturases.

Isolation of a full-length cDNA

To isolate a full-length cDNA clone, the 520 bp PCR fragment was DIG-labelled, and used to screen a cDNA library of 12-day-old protonemta. Of 3.0×10^5 plaques screened, 19 positives were isolated. The restriction analysis of their inserts showed a similar pattern in all cases. The partial sequence analysis from six inserts revealed that they were identical to each other within their overlapping regions and also to the original 520 bp PCR fragment. The longest insert, designated *PPDES6* cDNA, was sequenced on both strands. It had a length of 2012 bp excluding its poly(A) tail. An open reading frame stretched from position 319–1894, and several stop codons in the corresponding 5′ untranslated region indicated its full length (Figure 1). The protein PPDES6 translated from the *PPDES6* cDNA contained 525 amino acid residues with a calculated molecular weight of 59.3 kDa. This is 7–20 kDa larger than all acyl-lipid desaturases known from higher plants and cyanobacteria. Data bank searches indicated similarity to cytochrome b_5 sequences from residues 105–176 and to desaturases from residue 207 towards the C-terminus.

The desaturase domain showed the highest similarity to the cytochrome b_5-containing fusion protein of *Helianthus annuus* (Sperling *et al.*, 1995), a putative fusion protein from *Caenorhabditis elegans* encoded by cosmid T13F2 (Z81122) and the Δ6-desaturases of *Spirulina platensis* (X87094), *Borago officinalis* (Sayanova *et al.*, 1997) as well as *Synechocystis* sp. PCC 6803 (Reddy *et al.*, 1993). The identity values of PPDES6 to these proteins were low and ranged from 21% to 27% for the sequence between the first and third histidine boxes and from 12% to 23% over the entire length. The sequence motive QIEHH of the third histidine box started with a glutamine instead of a histidine, which has also been found in Δ6-desaturases and the cytochrome b_5 fusion protein of *H. annuus*, but not in other membrane-bound desaturases. The hydrophobicity plot (Kyte and Doolittle, 1982) after residue 200 showed the typical profile of membrane-bound desaturases (data not shown). The cytochrome b_5-related domain contained the eight invariant residues typical for the cytochrome b_5 superfamily (Lederer, 1994).

The N-terminal extension of about 100 residues did not share significant similarity to any sequence in the data banks, and computer analysis did not detect any motives for protein targeting or modification either for the extension or for the whole protein.

Structure of the gene

To knock out the *PPDES6* gene, its genomic sequence was amplified by PCR with specific primers C and D. Primer C was deduced from the 5′ end and D from the middle of the 3′ untranslated region of the *PPDES6* cDNA. PCR with these primers and genomic DNA of *P. patens* as template amplified a fragment that was 1578 bp longer than the distance between the binding sites of the primers on the cDNA. The genomic PCR fragment, denoted *PPDES6*, was cloned and sequenced on both strands (Figure 2). Apart from six putative introns (i1-i6) it was 100% identical with the cDNA, confirming its identity as the genomic locus of the *PPDES6* cDNA. The 5′ splicing border of five introns was GT and the 3′ border of all six was AG. Only the fourth intron i4 contained the unusual 5′ splicing border GC, which has been found in genes of several plant species (Xue and Rask, 1995). The reliability of this intron sequence was confirmed by sequencing two other PCR-amplified clones over this region. The intron i4 was located between two triplets coding for residues 176 and 177. After residue 176 the detected similarity to cytochrome b_5 sequences was terminated.

Gene targeting

For the disruption experiments, the first histidine box of the genomic clone was replaced by the *npt* II gene as a positive selection marker. The subsequent double digestion

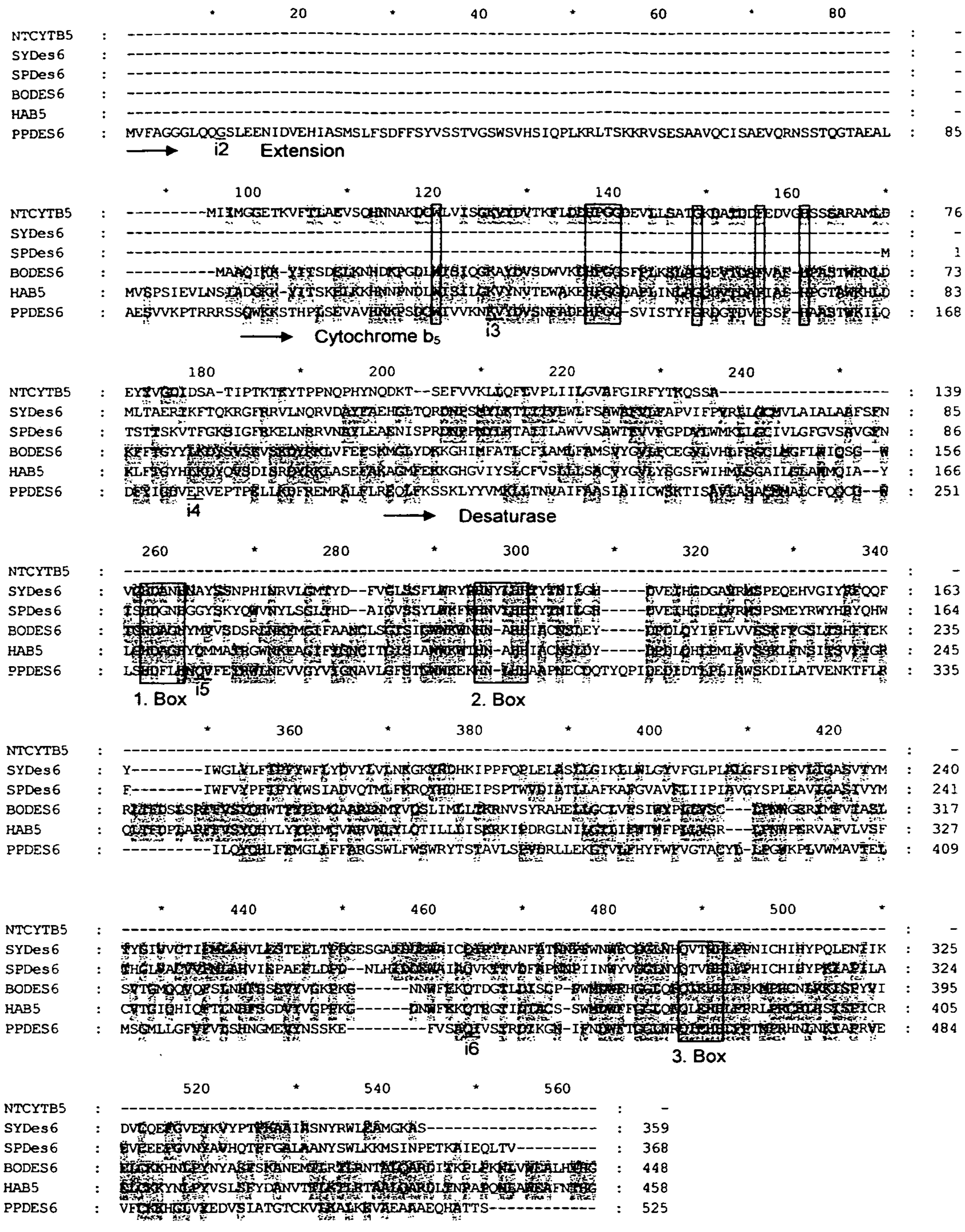


Figure 1. Amino acid sequences of PPDES6 and closely related proteins.
For alignment the CLUSTAL X program was used (gap opening 10, gap extension 0.05). Conserved and invariant residues are grey. The approximate beginning of the three domains from PPDES6 are marked by arrows and their putative function. The eight invariant residues characteristic for the cytochrome b_5 superfamily and the three histidine boxes of the desaturase domains are framed. The underlined residues indicate the positions of introns i1–i6 in the genomic sequence *PPDES6*. SYDes6, SPDes6 and BODES6 refer to the Δ6-desaturases of *Synechocystis* (U79010), *Spirulina* (X87094) and *Borago* (U79010). NTCYTB5 and HAB5 refer to the cytochrome b_5 of *Nicotiana* (X71441) and the b_5 fusion protein of *Helianthus* (X87143), respectively.

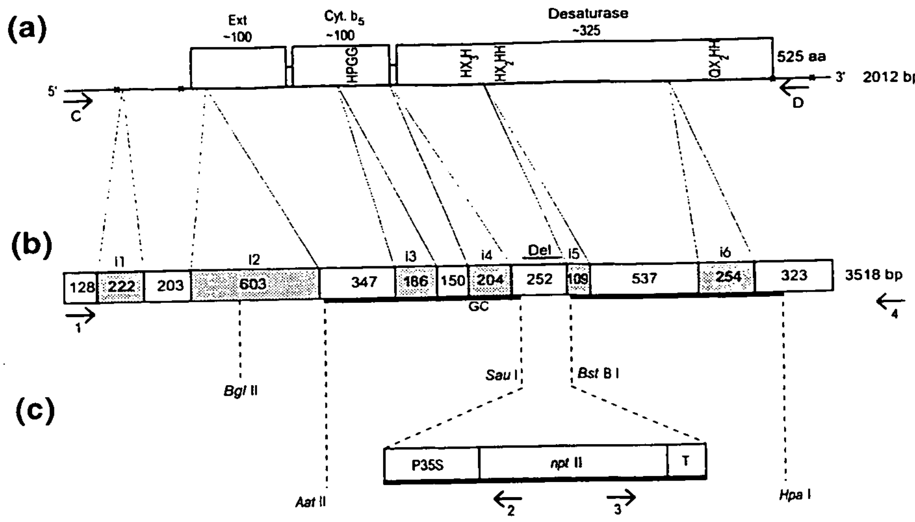


Figure 2. Structure of the desaturase cDNA (a), the desaturase gene (b) and the fragment used for gene disruption (c).
(a) *PPDES6* cDNA (2012 bp) and schematic representation of the protein PPDES6 (525 aa) with three domains: an extension (Ext) of unknown function, a cytochrome b_5- and a desaturase-like region with typical amino acid blocks. Stop codons are symbolized by stars.
(b) Corresponding genomic sequence *PPDES6* amplified with primer C and D (shown in a). The positions and lengths of the six introns (i1–i6, grey) are marked. The GC splice junction is shown for i4.
(c) Replacement fragment for gene disruption: the first histidine box in the genomic fragment was substituted after a *SauI*/*BstBI* double digestion by the *npt* II cartridge, which contained the *npt* II coding region between the CaMV 35S promoter (P35S) and terminator (T). The transformation was carried out after linearization with *AatII* and *HpaI* resulting in a linear fragment (underlined in bold) with the *npt* II cartridge inserted into the sequence of the desaturase. The numbered arrows below the blocks indicate the binding sites of primers used for subsequent PCR analyses. The localization of the Del probe used for Southern blotting (see Figure 3) is marked with a line above the block.

with *SauI*/*BstBI* yielded a linear fragment with the *npt* II gene in its centre and the desaturase arms at both ends (Figure 2). This linear fragment was used to transform *P. patens* protoplasts by the PEG method (Schaefer *et al.*, 1991). Seven transformation experiments with 3.0×10^5 protoplasts in each experiment resulted in the isolation of 56 independent and stably transformed lines. Five randomly selected transgenic lines (K1–K5) were used for detailed analysis regarding the molecular biology of gene disruption as well as its consequences for fatty acid biosynthesis.

Molecular analysis of the transgenic lines

The specific integration of the transformed DNA into the *PPDES6* gene was analysed by PCR using genomic DNA from five transformed lines (K1–K5) and the wild type. The locations of the different primers are presented in Figure 2. It is important to point out that the 3′ end of primer 4 binds 40 bp downstream of the cloned genomic sequence to exclude PCR signals resulting from contamination by the DNA used for transformation. Its sequence was derived from the 3′ end of an incomplete cDNA clone, which showed the same sequence in the overlapping region with cDNA *PPDES6*, but contained a longer 3′ end.

PCR with the primer pair 1/2 amplified fragments of 2.7 kbp, and with the primer pair 3/4 bands of 1.6 kbp, from all five transformants, whereas experiments with the wild type gave negative results. The length of the bands agreed with a substitution of the first histidine box of the *PPDES6* gene by the *npt* II cassette. Both PCR fragments from two transformants (K2 and K3) were cloned and partially sequenced. The sequenced segments were identical with the corresponding regions of the transformed gene disruption construct. Most important, the fragments from primer pair 3/4 contained the downstream genomic element of 40 bp, which was absent in the transformed DNA. They lacked the first histidine box, and the transition regions of the *npt* II cassette to the *PPDES6* gene, as well as the regions containing the restriction sites *Aat* II and *Hpa* I, were identical in their sequence with the disruption construct.

To provide evidence for a deletion of the first histidine box in the *PPDES6* gene of the transgenic lines, the genomic DNA of the transformed lines and the wild type was digested with *BglII*, blotted and hybridized with the DIG-labelled deletion probe Del. This probe represents the *SauI*/*BstBI* fragment encoding the first histidine box, which had been deleted from the transformed disruption construct (Figure 3). Hybridization with the deletion probe Del showed one strong signal of 4.5 kbp and two very weak signals of 5.0 and 7.0 kbp with the wild type DNA. The transformed lines K1–K4 had lost the strong 4.5 kbp signal but not the two weak signals. Line K5 corresponded to the

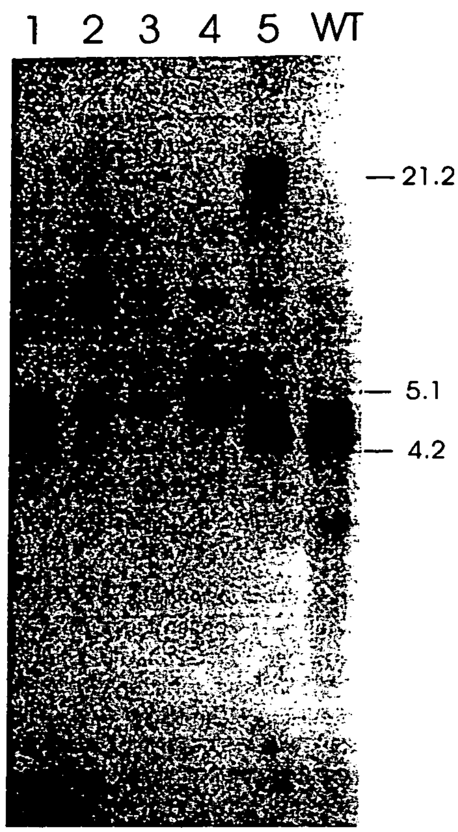


Figure 3. Verification of gene disruption by Southern blotting.
Genomic DNA (4 μg) from the wild type (WT) and five transformed lines K1–K5 (1–5) was digested with *Bgl*II and hybridized with the deletion probe (Del). The location of the probe is described in Figure 2. Molecular weights in kbp are indicated on the right.

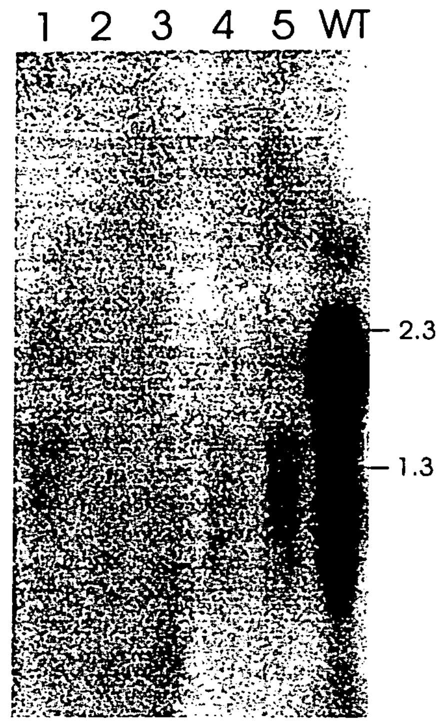


Figure 4. Northern blot analysis of *PPDES6* expression.
Total RNA (20 μg) from 14-day-old *P. patens* protonemata was probed with an RNA probe transcribed from the last 600 bp of the *PPDES6* cDNA. Five transgenic lines K1–K5 (1–5) and the wild type (WT) were analysed. Molecular weights in kb are indicated on the right.

wild type situation but contained an additional band of more than 21 kbp.

To compare the expression of *PPDES6* in the five transgenic lines with the wild type, we blotted total RNA of 14-day-old protonemata and hybridized it with a DIG-labelled RNA probe against the 3′ end of the *PPDES6* cDNA (Figure 4). The wild type showed a strong signal of 2.0–2.2 kb, whereas the five transgenic lines had lost this transcript. Hybridization with a *npt* II-specific probe (blot not shown) detected a strong signal of 1.0–1.3 kb in all transgenic lines but not in the wild type.

Functional analysis of PPDES6 in P. patens

For the functional identification of the desaturase, we analysed the total fatty acids of the wild type and the five knockout lines. The fatty acid analyses presented in Figure 5 are confined to the wild type and to line K2, but the other four lines tested gave essentially the same results. Pathways [1] and [2] below show the sequences proposed for the biosynthesis of AA (20:4) and EPA (20:5) in *P. patens*, and they are supported by our results (fatty acids are indicated as $m{:}n^{\Delta a,b,c,\ldots}$; m refers to the number of carbon atoms, n to the double bonds and $^{\Delta a,b,c,\ldots}$ to the position of the double bonds; desaturation and elongation steps are indicated by Δx and EL).

$$18{:}2^{\Delta 9,12} \xrightarrow{\Delta 6} 18{:}3^{\Delta 6,9,12} \xrightarrow{EL} 20{:}3^{\Delta 8,11,14} \xrightarrow{\Delta 5} 20{:}4^{\Delta 5,8,11,14} \quad [1]$$

$$\downarrow \Delta 15$$

$$18{:}2^{\Delta 9,12,15} \xrightarrow{\Delta 6} 18{:}4^{\Delta 6,9,12,15} \xrightarrow{EL} 20{:}4^{\Delta 8,11,14,17} \xrightarrow{\Delta 5} 20{:}5^{\Delta 5,8,11,14,17} \quad [2]$$

Compared with the wild type, all transgenic lines showed a strong decrease in those unsaturated fatty acids, the formation of which involves a Δ6-desaturation step (Figure 5): $18{:}3^{\Delta 6,9,12}$, $18{:}4^{\Delta 6,9,12,15}$, $20{:}3^{\Delta 8,11,14}$, $20{:}5^{\Delta 5,8,11,14,17}$ and most clearly $20{:}4^{\Delta 5,8,11,14}$. On the other hand, the possible substrates for a Δ6-desaturase, $18{:}2^{\Delta 9,12}$ and $18{:}3^{\Delta 9,12,15}$, increased. Therefore, it is most likely that the reactions from $18{:}2^{\Delta 9,12}$ to $18{:}3^{\Delta 6,9,12}$ as well as from $18{:}3^{\Delta 9,12,15}$ to $18{:}4^{\Delta 6,9,12,15}$ were blocked, both of which are catalysed by a Δ6-desaturase (compare pathways [1] and [2]).

To provide further evidence for the function of the new

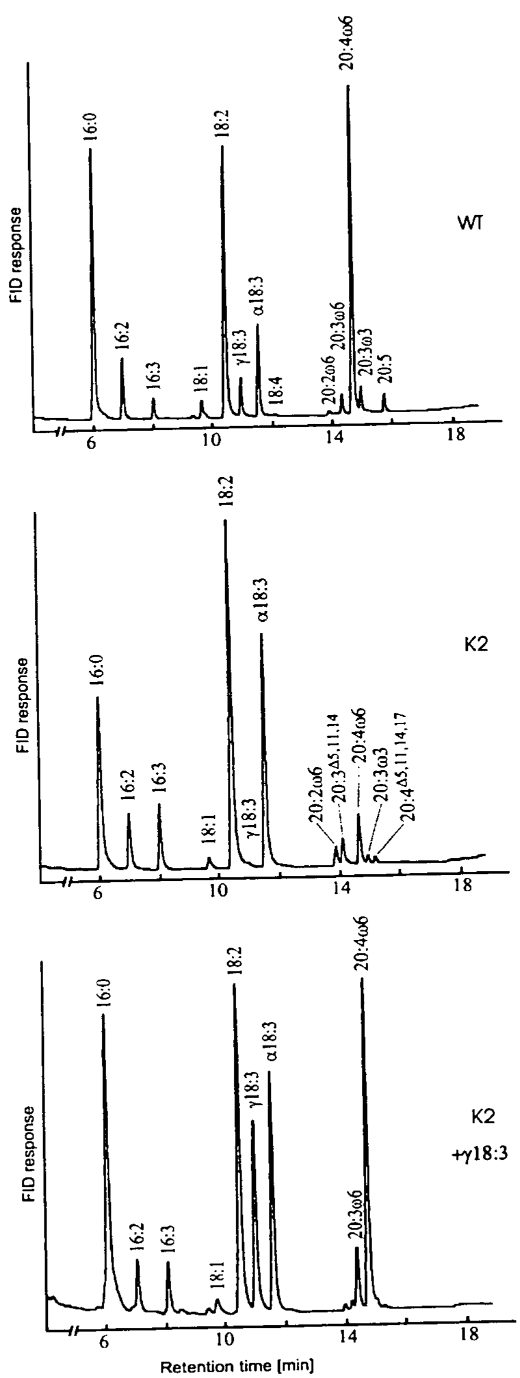


Figure 5. Fatty acid profiles of the *P. patens* wild type (WT) and the knockout line K2. The fatty acid methyl esters (FAME) of the total lipids were analysed by capillary gas-liquid chromatography. The chromatograms WT and K2 show the FAME of protonemata grown for 14 days in liquid medium. The lower chromatogram shows the FAME profile of K2 cells cultured under the same conditions but in the presence of 50 μM of γ18:3 ($18{:}3^{\Delta6,9,12}$).

Δ6-desaturase, we supplemented the knockout line K2 and the wild type with $18{:}3^{\Delta6,9,12}$ (γ18:3). In K2 the feeding of this fatty acid resulted in the reappearance of $20{:}3^{\Delta8,11,14}$ and $20{:}4^{\Delta5,8,11,14}$, whereas almost no change was observed in the wild type. This experiment indicates that the knockout line K2 is able to synthesize 20:4 from added $18{:}3^{\Delta6,9,12}$, but not from $18{:}2^{\Delta9,12}$, which increases in unsupplemented K2. However, the addition of $18{:}3^{\Delta6,9,12}$ did not result in a complementation of the almost complete disappearance of $20{:}5^{\Delta5,8,11,14,17}$ in K2.

The addition of $20{:}2^{\Delta11,14}$ and $20{:}3^{\Delta11,14,17}$ (data not shown) did not result in an increase of 20:4 and 20:5 in the wild type or in K2. Another interesting effect of the knockout was the completely different proportion of C20-fatty acids in K2 (7%) compared to the wild type (30%).

Functional expression of PPDES6 in Saccharomyces cerevisiae

To exclude the possibility that the loss of a Δ6-desaturase in the knockout lines is a consequence of a regulatory difference between the *Physcomitrella* wild type and knockout lines, *PPDES6* was functionally expressed in *Saccharomyces cerevisiae*. Plasmid pYESΔ6 containing the open reading frame of the *PPDES6* cDNA was transformed into the *S. cerevisiae* strain INVSC1. One clone transformed with pYESΔ6 and another with the empty vector pYES2 as control were grown for four to five generations after induction with 2% galactose in minimal medium. Since *S. cerevisiae* does not contain the dienoic fatty acid substrates required for a Δ6-desaturase, the expression was performed with supplementation of $18{:}2^{\Delta9,12}$ and $18{:}3^{\Delta9,12,15}$, respectively. In subsequent analyses of total fatty acids, the following Δ6-desaturated products were detected in the strain expressing PPDES6: $16{:}2^{\Delta6,9}$, $18{:}2^{\Delta6,9}$, $18{:}3^{\Delta6,9,12}$ and $18{:}4^{\Delta6,9,12,15}$ (Table 1). In the control cells, none of these fatty acids were detected. The production of these fatty acids with an additional Δ6-double bond confirmed that cDNA *PPDES6* encodes a Δ6-fatty acid desaturase.

Discussion

Structural properties

The cDNA and the genomic sequence *PPDES6* encoding a novel Δ6-desaturase from *P. patens* were cloned using a PCR-based approach. The deduced protein shared less than 27% identity with the recently cloned Δ6-desaturase from *B. officinalis* and with the Δ6-desaturases from cyanobacteria (Reddy *et al.*, 1993; Sayanova *et al.*, 1997). This is a surprisingly low value, as until now all desaturases of the same regioselectivity and the same subcellular compartment have been more highly conserved, even between distantly related organisms. For example, six

Table 1. Expression of the Δ6-desaturase in *S. cerevisiae*. The fatty acid methyl esters of the total lipids from cells transformed with pYES2 (WT control) and pYESΔ6 (Δ6-desaturase of *P. patens*) were analysed by GLC. The cells were cultured in minimal medium supplemented with 2% galactose for 24 h at 30°C. The last two columns show data from cultures supplemented with $18{:}2^{\Delta 9,12}$ (18:2) and $18{:}3^{\Delta 9,12,15}$ (α18:3)

	% total fatty acids			
	pYES2	pYESΔ6		
Fatty acids	–	–	+ 18:2	+ α18:3
16:0	16.4	16.1	23.8	25.8
$16{:}1^{\Delta 9}$	54.0	55.5	38.1	31.4
$16{:}2^{\Delta 6,9}$	–	4.2	1.7	–
18:0	3.2	2.4	4.0	4.7
$18{:}1^{\Delta 9}$	24.9	19.7	19.1	19.2
$18{:}2^{\Delta 6,9}$	–	0.6	0.2	–
$18{:}2^{\Delta 9,12}$	–	–	8.5	–
$18{:}3^{\Delta 6,9,12}$	–	–	4.0	–
$18{:}3^{\Delta 9,12,15}$	–	–	–	11.7
$18{:}4^{\Delta 6,9,12,15}$	–	–	–	3.0

other PCR fragments from *P. patens*, isolated in this screening, coded for putative Δ12- and Δ15-desaturases and displayed more than 60% identity to the corresponding desaturases of higher plants and cyanobacteria.

The presence of the cytochrome b_5-related domain upstream of the desaturase suggests its localization in microsomes rather than in chloroplasts, because plastidial desaturases normally use ferredoxin as electron donor (Heinz, 1993). Besides this, PPDES6 contains a new N-terminal extension of about 100 amino acids, which is absent in other presently known desaturases. The function of this extension is unclear, since it shows no significant homology to any known protein, and targeting or modification signals were not detected. Interestingly, the three histidine boxes and the cytochrome b_5 domain of PPDES6 are encoded by separate exons (Figure 2), implying that they may constitute separate evolutionary units. The fourth intron containing the unusual 5′ splicing border GC is located directly after the last triplet for the cytochrome b_5 domain. This organization could allow a differential splicing between the 5′ border of the first and the 3′ border of the fourth intron, resulting in a deletion of both the cytochrome b_5 domain and the N-terminal extension from the desaturase domain of the *PPDES6* transcript.

Molecular analysis of the transgenic lines

In this study, we have described the highly efficient knockout of the *PPDES6* gene after transforming *P. patens* with a linear disruption fragment. PCR experiments proved the specific integration of the *npt* II cassette into the *PPDES6* locus in all arbitrarily chosen transgenic lines. Furthermore, Southern blot experiments confirmed the deletion of a 200 bp segment encoding the first histidine box from the genome of four transgenic lines (K1–K4). It is likely that reciprocal exchange by double cross-over led to the integration observed in these four lines. Targeting experiments from Schaefer and Zrÿd (1997) demonstrated homologous integration into a locus but not a substitution. The blots with line K5 reveal an even more complicated situation. Nevertheless, K5 does not express the Δ6-desaturase activity any more. Two additional signals of low intensity in wild type and in all transgenic lines indicated that related genomic sequences were not involved in the gene targeting events. The presence of these sequences suggests that isoforms of other Δ6-desaturases could be expressed to some extent in the knockout lines.

In the Northern blots all transgenic lines showed a dramatically reduced expression of *PPDES6* while this transcript was abundant in the wild type. Thus loss of desaturase activity, as evident from the fatty acid profiles most probably resulted from loss of transcription due to gene disruption.

Functional analysis of PPDES6 in P. patens *and* S. cerevisiae

The gene disruption of *PPDES6* resulted in a dramatic alteration of the fatty acid pattern in the transformed lines. The knockout lines showed an increase of 18:2 and α18:3 and a decrease of Δ6-desaturated fatty acids. Therefore, it is likely that *PPDES6* codes for a Δ6-desaturase, which desaturates $18{:}2^{\Delta 9,12}$ to $18{:}3^{\Delta 6,9,12}$ and $18{:}3^{\Delta 9,12,15}$ to $18{:}4^{\Delta 6,9,12,15}$. The Δ6-regioselectivity of PPDES6 was further verified by restoration of 20:4 biosynthesis upon feeding of γ18:3 (Figure 5). The synthesis of 20:4 from γ18:3 would not work if a Δ5-desaturase or the elongation system had been blocked. The Δ6-desaturation of 18:2 and α18:3 added to *S. cerevisiae* cells expressing PPDES6 confirmed these results and excluded the possibility that the loss of a Δ6-desaturase in the knockout lines was due to regulatory alterations, for example the loss of an activator for the Δ6-desaturase. On the other hand, we could not detect a Δ8-C20-desaturase in *P. patens*, since addition of $20{:}2^{\Delta 11,14}$ and $20{:}3^{\Delta 11,14,17}$ did not increase the content of 20:4 and 20:5. A Δ8-desaturase operating at the C20-level could theoretically replace the Δ6-C18-desaturase in the biosynthesis of 20:4 and 20:5. Such an enzyme has been suggested to be present in *Euglena gracilis* (Nichols and Appleby, 1969).

Based on the knockout effects and feeding experiments, we propose the two pathways [1] and [2] mentioned above for the biosynthesis of 20:4 and 20:5 in *P. patens*, which branch at 18:2. They are in agreement with the biosynthesis of 20:4 and 20:5 as suggested for *Porphyridium cruentum* (Shiran *et al.*, 1996).

It should be noted that *S. cerevisiae* cells expressing PPDES6 produced not only $18{:}3^{\Delta6,9,12}$ and $18{:}4^{\Delta6,9,12,15}$, but also $16{:}2^{\Delta6,9}$ and $18{:}2^{\Delta6,9}$, which were not detected in *P. patens*. The reason for their absence in *P. patens* may be the low content and rapid turn-over of the putative precursors, $16{:}1^{\Delta9}$ and $18{:}1^{\Delta9}$, in the moss, whereas they are produced in high amounts by *S. cerevisiae*. Since the Δ6-desaturase converts $16{:}1^{\Delta9}$ to $16{:}2^{\Delta6,9}$, but does not introduce a Δ8-double bond into $20{:}2^{\Delta11,14}$ and $20{:}3^{\Delta11,14,17}$ (mentioned above), the insertion of the Δ6-double bond involves measuring from the carboxy terminus (and the Δ9-double bond) rather than from the methyl end. This classifies the desaturase as a Δ6-desaturase (Heinz, 1993).

Another interesting effect is the significant decrease in C20-fatty acids in the knockout lines. The decrease from more than 30% in the wild type to less than 7% in K2 indicates that the elongation system of *P. patens* prefers or even requires Δ6-desaturated C18-fatty acids. This elongation process is either very rapid or channelled and thus prevents the accumulation of γ18:3 or 18:4 in lipids. In the other organisms, from which Δ6-desaturases have been cloned (*B. officinalis* and *Synechocystis*), elongation systems do not co-operate with this desaturase and therefore Δ6-desaturated fatty acids can accumulate. A detailed analysis of lipids and fatty acids in *P. patens* wild type and knockout plants, as well as in *S. cerevisiae* expressing the Δ6-desaturase, will be published elsewhere (T. Girke *et al.*, manuscript in preparation).

In our present study, all knockout lines still contained small amounts of fatty acids, which were synthesized by a pathway requiring Δ6-desaturase. This indicates that at least one other functional gene for a Δ6-desaturase should exist. Possible candidates may be the two faint signals observed above the targeted 4.5 kbp fragment in Southern blots of wild type and transgenic lines (Figure 3).

Apart from these biochemical changes, we did not detect any visibly altered phenotype in the knockout plants, at least in their protonema or gametophore states at 25°C. Therefore, it was not possible at this point to evaluate the physiological importance of 20:4 for the moss. The appearance of a visible phenotype may also be prevented by residual 20:4. Deletions of several desaturases in *Synechocystis* became critical only if the Δ6- and Δ12-desaturase were knocked out together, whereas a reduction in trienoic acids without affecting dienoic acids was not critical (Tasaka *et al.*, 1996).

Experimental procedures

Plant material and culture conditions

The protonemata of *Physcomitrella patens* (Hedw.) BSG were grown in liquid medium (Reski *et al.*, 1994). For feeding experiments with fatty acids, 4-day-old cultures were supplemented with ammonium salts of fatty acids (dissolved in ethanol) to a final concentration of 50 μM and further cultivated for an additional 6–8 days.

Analysis of nucleic acids

DNA manipulations were performed according to standard protocols (Sambrook *et al.*, 1989) unless otherwise stated. DNA sequences were determined on both strands by the dideoxy chain termination method using Dye Primer as well as Dye Terminator sequencing kits.

PCR with degenerated primers and cDNA library screening

Poly(A)$^+$ RNA was isolated with Dynabeads (Dynal, Oslo, Norway) from total RNA of 12-day-old *P. patens* protonema cultures, and reverse-transcribed into single-stranded cDNA. This ss-cDNA was used as template in the PCR-based cloning. A 550 bp PCR fragment was amplified with the degenerate sense primer A 5′-TGGTGGAA(A/G)TGGA(C/A)ICA(T/C)AA-3′ and antisense primer B 5′-GG(A/G)AA(A/T/G/C)A(A/G)(G/A)TG(G/A)TG(C/T)TC-3′ derived from the amino acid sequence WWKW (N/T)HN and EHHLFP, respectively. The PCR reactions were carried out with Taq DNA polymerase using an amplification programme of 3 min denaturation at 94°C, followed by 30 cycles of 20 sec at 94°C, 30 sec at 45°C, 1 min at 72°C and terminated by 5 min extension at 72°C. The PCR fragments of the expected length (500–600 bp) were cloned in pUC18 and sequenced. A digoxygenin-labelled DNA probe of the PCR fragment was synthesized by PCR and used to screen a lambda ZAPII cDNA library of 12-day-old protonemata according to the manufacturer's protocols (Boehringer, Mannheim, Germany; Stratagene, La Jolla, CA). The longest insert (*PPDES6* cDNA) was sequenced on both strands using overlapping subclones. The corresponding genomic sequence *PPDES6* was isolated by PCR with specific primers C (5′-CCGAGTCGCGGATCAGCC-3′) and D (5′-CAGTACATTCGGTCATTCACC-3′) using the Expand High Fidelity PCR System (Boehringer) and the hot start PCR program described below. *PPDES6* was cloned into the pCR-Script Amp SK(+) cloning vector (Stratagene), resulting in plasmid p*PPDES6* and sequenced on both strands.

Transformation of P. patens

First the vector pRT101neo was constructed to obtain a *npt* II selection cassette, which could be excised by *Hin*dIII digestion. For this purpose the *npt* II coding region of pRT100neo (Töpfer *et al.*, 1993) was excised with *Hin*dIII (blunted)/*Xho*I and ligated between the CaMV 35S promoter and terminator of pRT101 (Töpfer *et al.*, 1987), which had been digested with *Xba*I (blunted)/*Xho*I. The gene disruption construct resulted from the substitution of a *Sau*I/*Bst*BI fragment in the genomic clone p*PPDES6* by the *npt* II selection cartridge. Subsequently, the disruption construct was digested with *Aat*II and *Hpa*I, resulting in a linear fragment with the *npt* II gene in its centre flanked by genomic sequences of 923 bp and 1159 bp. Fifteen micrograms of this linear DNA were phenol extracted, precipitated and used for the transformation without separation from the vector. PEG-mediated direct DNA transfer into protoplasts was performed as described by Schaefer *et al.* (1991). The regenerated protonemata were selected for 14 days on medium with G418 (50 mg l^{-1}), released for 12 days under non-selective conditions and again grown for 14 days on

selection plates. Well growing plants surviving this selection regime were defined as stable transformants and cultivated for mass production in non-selective liquid medium. The stability of their G418 resistance was tested every 4 weeks by incubating aliquots on selection plates.

DIG-labelling of DNA and RNA

DNA probes were labelled with digoxygenin by PCR (PCR DIG probe synthesis kit; Boehringer). The 5′ ends of the primers for the deletion probe (Del) were located on the *PPDES6* cDNA at position 910 and 1092 (*SauI*/*BstBI* fragment). The desaturase RNA probe was transcribed by *in vitro* transcription with digoxygenin (Boehringer) from a subclone of the *PPDES6* cDNA containing the last 600 bp of its 3′ end and the *npt* II probe from a subclone coding for the *npt* II.

PCR detection, Southern and Northern blot analysis

Four primers were used in the PCR experiments for the detection of gene targeting events. Primer 1 was derived from the 5′ end of *PPDES6*. Primers 2 and 3 were constructed from the ends of the *npt* II coding region. The sequence of primer 4 (5′-CAGAGACGAATCGTGGCTCC-3′) was derived from the 3′ end of an incomplete cDNA clone, which was identical with *PPDES6* cDNA in the overlapping region, but contained a longer 3′ end. The PCR experiments with these primers were run with a hot start programme of 10 min denaturation at 94°C, addition of the polymerase at 72°C, followed by 30 cycles of 30 sec at 94°C, 30 sec at 60°C, 3 min at 72°C and terminated by 10 min extension at 72°C. Genomic DNA of *P. patens* was extracted with cetyl-trimethyl-ammonium bromide according to Rogers and Bendich (1988). Four micrograms of DNA were digested with the appropriate restriction enzyme, separated on a 0.7% agarose gel by electrophoresis, transferred onto a nylon membrane and hybridized. The final washing steps were performed in 0.5 × SSC with 0.1% SDS at 68°C. The detection was accomplished with a chemiluminescent substrate (CSPD, Boehringer). The Northern blot experiments were performed with total RNA isolated from 14-day-old protonema cultures (RNeasy plant kit, Qiagen, Hilden, Germany). Twenty micrograms of total RNA were separated on a standard formaldehyde gel, blotted onto a nylon membrane and hybridized with RNA probes. The final washing steps were performed in 0.1 × SSC with 0.1% SDS.

Expression in S. cerevisiae

The open reading frame of the *PPDES6* cDNA was cloned behind the galactose-inducible promotor *GAL1* of the yeast expression vector pYES2 (Invitrogen, Leek, Netherlands). For this purpose, a new *XhoI* site was introduced by PCR (32 bp upstream from its deduced translational start at position 319). The entire open reading frame of the desaturase was released with *HindIII* (blunted)/*XhoI* and ligated into the *XbaI* (blunted)/*XhoI* sites of the pYES2 vector to yield plasmid pYESΔ6. Its sequence was verified by DNA sequencing. The plasmids pYESΔ6 and pYES2 were transformed into the *Saccharomyces cerevisiae* strain INVSC1 (Invitrogen) by the lithium acetate method (Ausubel *et al.*, 1995). Cells harbouring the plasmids pYES2 and pYESΔ6 were grown in complete minimal drop-out uracil medium (CMdum) containing 2% raffinose as the exclusive carbon source (Ausubel *et al.*, 1995; Kajiwara *et al.*, 1996) and 1% Tergitol NP-40 (w/v; Sigma) for the solubilization of fatty acids (Avery *et al.*, 1996). For expression experiments, the cultures were grown to an optical density (600 nm) of 0.5 in CMdum medium, then supplemented with 2% galactose (w/v) as well as 0.003% of the corresponding fatty acid (w/v; stock solution solubilized in 5% tergitol) and finally grown to saturation for 24 h at 30°C.

Lipid analysis

Lipids were extracted from protonemata and yeast cells by chloroform-methanol extraction (Siebertz *et al.*, 1979) and purified from apolar components by TLC in diethylether. In this solvent all membrane lipids (triacylglycerols were not produced by protonemata) remained at the start. The fatty acid methyl esters (FAME) were obtained by transmethylation of the lipids with 1 N H_2SO_4 in methanol and 2% dimethoxypropane at 80°C for 1 h. The extracted FAME were analysed by gas-liquid chromatography using a capillary column (Chrompack, WCOT Fused Silica, CP-Wax-52 CB, 25 m, 0.32 mm). Their identities were confirmed by comparison with appropriate FAME standards (Sigma). The corresponding fatty acid pyrrolidides were obtained as described elsewhere (Andersson and Holman, 1974) and analysed by GLC-MS on a HP 5989 A instrument (Hewlett-Packard) equipped with an HP-5 column using a temperature gradient 150°C (3 min) → 320°C at 5° min^{-1}. Electron impact (EI) was carried out at 70 eV and chemical ionization mass spectra (CI-MS) were recorded with ammonia as reactant gas (0.1 MPa).

Acknowledgements

This work is based on a doctoral study by Thomas Girke in the Faculty of Biology, University of Hamburg. Support by the Deutsche Forschungsgemeinschaft (He 695/14–1) and the Fonds der Chemischen Industrie is gratefully acknowledged. Ralf Reski is indebted to the Deutsche Forschungsgemeinschaft for a Heisenberg Fellowship (Re 837/3–1). We thank Hermann Moll for skilful help in GC–MS analysis and our colleagues in the EC-funded project EUROMOSS, who shared unpublished results.

References

Andersson, B.Å. and Holman, R.T. (1974) Pyrrolidides for mass spectrometric determination of the position of the double bond in monounsaturated fatty acids. *Lipids*, **9**, 185–190.

Arao, T. and Yamada, M. (1994) Biosynthesis of polyunsaturated fatty acids in the marine diatom, *Phaeodactylum tricornutum*. *Phytochemistry*, **35**, 1177–1181.

Ashton, N.W. and Cove, D.J. (1977) The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants in the moss *Physcomitrella patens*. *Mol. Gen. Genet.* **154**, 87–95.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K., Albright, L.M., Coen, D.M. and Varki, A. (1995) *Current Protocols in Molecular Biology*. New York: John Wiley & Sons.

Avelange-Macherel, M.-H., Macherel, D., Wada, H. and Murata, N. (1995) Site-directed mutagenesis of histidine residues in the Δ12 acyl-lipid desaturase of *Synechocystsis*. *FEBS Lett.* **361**, 111–114.

Avery, S.V., Howlett, N.G. and Radice, S. (1996) Copper toxicity towards *Saccharomyces cerevisiae*: dependence on plasma membrane fatty acid composition. *Appl. Environ. Microbiol.* **62**, 3960–3966.

Cohen, Z., Norman, H.A. and Heimer, Y.M. (1995) Microalgae as a source of ω3 fatty acids. In *Plants in Human Nutrition*

(Simopoulos, A.P., ed.). World Rev. Nutr. Diet., Vol. 77. Basel: Karger, pp. 1–31.

Dembitsky, V.M. (1993) Lipids of bryophytes. *Prog. Lipid. Res.* **32**, 281–356.

Enoch, H.G., Catalá, A. and Strittmatter, P. (1976) Mechanism of rat liver microsomal stearoyl-CoA desaturase. *J. Biol. Chem.* **251**, 5095–5103.

Grimsley, N.H., Grimsley, J.M. and Hartmann, E. (1981) Fatty acid compositions of mutants of the moss *Physcomitrella patens*. *Phytochemistry*, **20**, 1519–1524.

Heinz, E. (1993) Biosynthesis of polyunsaturated fatty acids. In *Lipid Metabolism in Plants* (Moore, T.S. Jr, ed.). Boca Raton, FL: CRC Press, pp. 33–89.

Jamieson, G.R. and Reid, E.H. (1975) The fatty acid composition of fern lipids. *Phytochemistry*, **14**, 2229–2232.

Kajiwara, S., Shirai, A., Fujii, T., Toguri, T., Nakamura, K. and Ohtaguchi, K. (1996) Polyunsaturated fatty acid biosynthesis in *Saccharomyces cerevisiae*: expression of ethanol tolerance and the *FAD2* gene from *Arabidopsis thaliana*. *Appl. Environ. Microbiol.* **62**, 4309–4313.

Kearns, E.V., Hugly, S. and Somerville, C. (1991) The role of cytochrome b_5 in Δ^{12} desaturation of oleic acid by microsomes of safflower (*Carthamus tinctorius* L.). *Arch. Biochem. Biophys.* **284**, 431–436.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132.

Lederer, F. (1994) The cytochrome b5-fold: an adaptable module. *Biochimie*, **76**, 674–692.

Nichols, B.W. and Appleby, R.S. (1969) The distribution and biosynthesis of arachidonic acid in algae. *Phytochemistry*, **8**, 1907–1915.

Ohlrogge, J.B., Jarworski, J.G. and Post-Breitenmüller, D. (1993) *De novo* fatty acid biosynthesis. In *Lipid Metabolism in Plants* (Moore, T.S. Jr, ed.). Boca Raton, FL: CRC Press, pp. 3–32.

Reddy, A.S., Nuccio, M.L., Gross, L.M. and Thomas, T.L. (1993) Isolation of a Δ^6-desaturase gene from the cyanobacterium *Synechocystis* sp. strain PCC 6803 by gain-of-function expression in *Anabaena* sp. strain PCC 7120. *Plant Mol. Biol.* **27**, 293–300.

Reski, R., Faust, M., Wang, X.H., Wehe, M. and Abel, W.O. (1994) Genome analysis of the moss *Physcomitrella patens* (Hedw.) BSG. *Mol. Gen. Genet.* **244**, 352–359.

Rogers, S.O. and Bendich, A.J. (1988) Extraction of DNA from plant tissue. In *Plant Molecular Biology Manual* (Gelvin, S.B., Schilperoort, R.A. & Verma, D.P.S., eds), Section A6. Dordrecht, the Netherlands: Kluwer Academic Publishers, pp. 1–10.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Samuelsson, B. (1983) Leucotrienes: mediators of immediate hypersensitivity and inflammation. *Science*, **220**, 568–575.

Sayanova, O., Smith, M.A., Lapinskas, P., Stobart, A.K., Dobson, G., Christie, W.W., Shewry, P.R. and Napier, J.A. (1997) Expression of a borage desaturase cDNA containing an N-terminal cytochrome b_5 domain results in the accumulation of high levels of Δ6-desaturated fatty acids in transgenic tobacco. *Proc. Natl Acad. Sci. USA*, **94**, 4211–4216.

Schaefer, D. and Zrÿd, J.-P. (1997) Efficient gene targeting in the moss *Physcomitrella patens*. *Plant J.* **11**, 1195–1206.

Schaefer, D., Zrÿd, J.-P., Knight, C.D. and Cove, D.J. (1991) Stable transformation of the moss *Physcomitrella patens*. *Mol. Gen. Genet.* **226**, 418–424.

Schmidt, H. and Heinz, E. (1990) Involvement of ferredoxin in desaturation of lipid-bound oleate in chloroplasts. *Plant Physiol.* **94**, 214–220.

Shanklin, J., Whittle, E. and Fox, B.G. (1994) Eight histidine residues are catalytically essential in a membrane-associated iron enzyme, stearoyl-CoA desaturase, and are conserved in alkane hydroxylase and xylene monooxygenase. *Biochemistry*, **33**, 12787–12794.

Shiran, D., Khozin, I., Heimer, Y.M. and Cohen, Z. (1996) Biosynthesis of eicosapentaenoic acid in the microalga *Porphyridium cruentum*. I. The use of externally supplied fatty acids. *Lipids*, **31**, 1277–1282.

Siebertz, H.P., Heinz, E., Linscheid, M., Joyard, J. and Douce, R. (1979) Characterization of lipids from chloroplast envelopes. *Eur. J. Biochem.* **101**, 429–438.

Smith, M.A., Cross, A.R., Jones, O.T.G., Griffiths, W.T., Stymne, S. and Stobart, K. (1990) Electron-transport components of the 1-acyl-2-oleoyl-*sn*-glycero-3-phosphocholine Δ^{12} desaturase (Δ^{12} desaturase) in microsomal preparations from developing safflower (*Carthamus tinctorius* L.) cotyledons. *Biochem. J.* **272**, 23–29.

Sperling, P., Schmidt, H. and Heinz, E. (1995) A cytochrome b_5 fusion protein similar to plant acyl lipid desaturases. *Eur. J. Biochem.* **232**, 798–805.

Tasaka, Y., Gombos, Z., Nishiyama, Y., Mohanty, P., Ohba, T., Ohki, K. and Murata, N. (1996) Targeted mutagenesis of acyl-lipid desaturases in *Synechocystis*: evidence for the important roles of polyunsaturated membrane lipids in growth, respiration and photosynthesis. *EMBO J.* **15**, 6416–6425.

Töpfer, R., Maas, C., Höricke-Grandpierre, C., Schell, J. and Steinbiss, H.-H. (1993) Expression vectors for high-level gene expression in dicotyledonous and monocotyledonous plants. *Methods Enzymol.* **217**, 66–78.

Töpfer, R., Matzeit, V., Gronenborn, B., Schell, J. and Steinbiss, H.-H. (1987) A set of plant expression vectors for transcriptional and translational fusions. *Nucl. Acids Res.* **15**, 5890.

Xue, J. and Rask, L. (1995) The unusual 5′ splicing border GC is used in myrosinase genes of Brassicaceae. *Plant Mol. Biol.* **29**, 167–171.

Zhukova, N. and Aizdaicher, N.A. (1995) Fatty acid composition of 15 species of marine microalgae. *Phytochemistry*, **39**, 351–356.

EMBL nucleotide sequence database accession numbers AJ222980 and AJ222981.

VOL 26 N° 1 Jan Feb 1988

Société Française de
Physiologie Végétale

ISSN 0981-9428
PPBIEX

PLANT PHYSIOLOGY AND BIOCHEMISTRY

Formerly **PHYSIOLOGIE VÉGÉTALE**

An official journal of the Federation of
European Societies of Plant Physiology

Published with the support of INRA

gauthier-villars

Plant Physiol. Biochem., 1988, **26** (6), 777-792.

Review

Biosynthesis and regulation of linolenic acid in higher plants

Xuemin M. Wang and David F. Hildebrand

Department of Agronomy, University of Kentucky, Lexington, Kentucky, USA.

(Received for publication December 28, 1987; accepted April 14, 1988)

Wang, X.-M. and Hildebrand D. F. 1988. Biosynthesis and regulation of linolenic acid in higher plants. — Plant Physiol. Biochem. 26: 777-792.

The biosynthesis of polyunsaturated fatty acids in higher plants is reviewed with particular emphasis on linolenate biosynthesis. Much information has been gained concerning linolenic acid synthesis by following the fate of radiolabelled precursers *in vivo*. Linolenate synthesis apparently occurs in both the chloroplasts on galactolipids and the endoplasmic reticulum on phospholipids. Linoleate desaturation can be differentially affected by chemical modulators and environmental conditions such as temperature, light and water stress relative to fatty acid biosynthesis resulting in changes in the linolenate content of lipids. Progress on the biochemical characterization of linoleoyl desaturase has been hampered by the apparent instability of the enzyme and the lack of a good *in vitro* assay system. Progress has been made in the breeding of plants for altered seed linolenate content (and other fatty acids) and a number of mutants have been found with altered linolenate levels of seed lipids and some of leaf lipids. Many of these mutants involve only one or two genes and therefore should be very useful in the biochemical and molecular characterization of linolenate biosynthesis in higher plants. The prospects for the genetic engineering of plants for altered fatty acid composition are discussed.

Additional key words — Lipids, fatty acids, polyunsaturated, oils, desaturation.

X.-M. Wang, Department of Botany, Louisiana State University, Baton Rouge, LA 70803 *USA; D. F. Hildebrand (reprint requests), Department of Agronomy, University of Kentucky, Lexington, KY* 40546, *USA.*

Résumé. *La biosynthèse des acides gras polyinsaturés chez les plantes supérieures est passée en revue en insistant spécialement sur la biosynthèse du linolénate. De nombreuses informations ont été obtenues sur la synthèse de l'acide linolénique en suivant* in vivo *le devenir de précurseurs marqués. Apparemment, la synthèse du linolénate a lieu à la fois dans les chloroplastes au niveau des galactolipides et dans le réticulum endoplasmique au niveau des phospholipides. La désaturation du linoléate peut être affectée de différentes manières par des agents chimiques ou des conditions de l'environnement, telles que la température, la lumière, le stress hydrique, qui agissent sur la biosynthèse des acides gras et dont le résultat est une modification du contenu en linolénate des lipides. Les progrès dans la caractérisation biochimique de la désaturase de l'acide linoléique ont été entravés par l'apparente instabilité de l'enzyme et l'absence d'un bon test d'activité* in vitro. *Des progrès ont été faits dans la sélection de plantes dont le contenu des graines en linolénate (et en autres acides gras) a été modifié, et des mutants présentant des teneurs modifiées en linolénate dans les lipides des graines et même des feuilles ont été obtenus. La plupart de ces mutations ne concernent qu'un ou deux gènes; elles devraient.*

Plant Physiol. Biochem., 0981-9428/88/06 777 16/$ 3.60 © Gauthier-Villars

donc être très utiles pour la caractérisation biochimique et moléculaire de la biosynthèse du linolénate chez les plantes supérieures. Les perspectives ouvertes pour la production de plantes ayant des compositions en acides gras modifiées sont discutées. Mots clés additionnels : lipides, acides gras, huiles polyinsaturées, désaturation.

Abbreviations. ACP, acyl carrier protein; Ch, choline; CoA, coenzyme A; DAG, diacylglycerol; CDPCh, cytidine diphosphoryl choline; EMS, ethylmethane sulfonate; ER, endoplasmic reticulum; G3P, glycerol 3-phosphate; LPA, lysophosphatidic acid; MGDG, monogalactosyldiacylglycerol; PA, phosphatidic acid; PC, phosphatidylcholine; PLTP, phospholipid transfer protein; TAG, triacylglycerol. C16:0, C18:3, etc., denote number of carbon atoms and double bonds. Pairs of numbers denoting fatty acids and separated by a slash (virgule) for example 18:2/18:3, represent the components at the *sn*-1 and *sn*-2 positions in PC, respectively. *Sn*-1 and *sn*-2 represent the first and second (middle) positions on the glycerol backbone of lipids.

Introduction

Photosynthetic tissues of higher plants typically contain 60-70% of linolenic acid (18:3) which is the most abundant fatty acid in nature (Gounaris and Barber, 1983). The presence of high levels of polyunsaturated fatty acids in plants has been implicated as playing a role in maintaining membrane fluidity of the photosynthetic apparatus and in preventing chilling damage (Raison, 1980; Oquist and Liljenberg, 1981; Harwood, 1983; Quinn and Williams, 1983; Kuiper, 1984).

Linolenic acid is a constituent of seed oil fatty acids in a number of oilseed crops such as soybean, rapeseed, and linseed. The quality of a seed oil is primarily dependent upon its fatty acid composition, which also determines its end use. Soybean oil contains 8% linolenic acid while other oilseed crops such as sesame, cottonseed, and sunflower contain less than 2%. The relatively high 18:3 level in soybean oil is not desirable for its use as a cooking oil due to its inverse correlation with oxidative stability and flavor quality (Smouse, 1979). Commercial soybean oil is a product of refining and industrial hydrogenation of the polyunsaturated fatty acids in the seed oil, which reduces the level of 18:3 and other unsaturated fatty acids. This expensive process also generates isomers of unsaturated fatty acids which are of concern in human health. Therefore, lowering the 18:3 content in soybean seed oil has been endeavored in several laboratories (Howell *et al.*, 1972; Trémolières *et al.*, 1978, 1982; Hammond and Fehr, 1983; Carver and Wilson, 1984; Wilcox *et al.*, 1984). However, progress in the development of commercial cultivars with lower 18:3 content has been slow.

A major factor contributing to the slow progress is the poor understanding of the biosynthesis and regulation of linolenic acid in both leaf and seed tissues (Stumpf, 1980; Frentzen, 1986). The formation of linolenic acid is considered to occur via consecutive desaturations of stearic, oleic, and linoleic acids with each step being catalyzed by a different enzyme (Trémolières and Mazliak, 1974; Cherif *et al.*, 1975; Slack *et al.*, 1978; Roughan *et al.*, 1979 *a;* Stymne and Stobart, 1985). However, little success has been achieved in the attempt to assay linoleoyl desaturase activity *in vitro*, which hampers further isolation and biochemical study of this enzyme. Further, it is not yet fully established which lipids are substrates for desaturation; how many distinct desaturases exist; whether 18:2 to 18:3 conversion occurs outside as well as inside the chloroplast; and how the formation of 18:3 is biochemically and genetically regulated. The present review discusses the current understanding of certain aspects of the biosynthesis and regulation of linolenate in higher plants.

Biosynthesis of polyunsaturated fatty acids

Desaturation of C18 fatty acids

The processes in the synthesis of plant polyunsaturated fatty acids from acetate are well understood up to the step of production of oleic acid (*fig.* 1). Enzymes involved in the formation of saturated fatty acids as far as stearic acid are all soluble, residing in the stromal phase of plastids (Stumpf, 1980).

The synthesis of oleic, linoleic, and linolenic acids in higher plants is thought to occur through consecutive desaturations from stearic acid (Roughan *et al.*, 1979 *a;* Jaworski, 1987). The first step of desaturation from stearate to oleate is catalyzed by stearoyl-ACP desaturase (Nagai and Bloch, 1968) and occurs in the stromal phase of chloroplasts. Unlike all other known desaturases, this desaturase is soluble instead of membrane bound and was the first plant desaturase to be studied in great detail in plants (Gurr, 1974). This enzyme uses stearoyl-ACP as the substrate and yields oleoyl-ACP as the product (Stumpf and Porra, 1976; Ohlrogge *et al.*, 1978). The reaction requires an NADH ferredoxin reductase, ferredoxin, as well as the desaturase; and

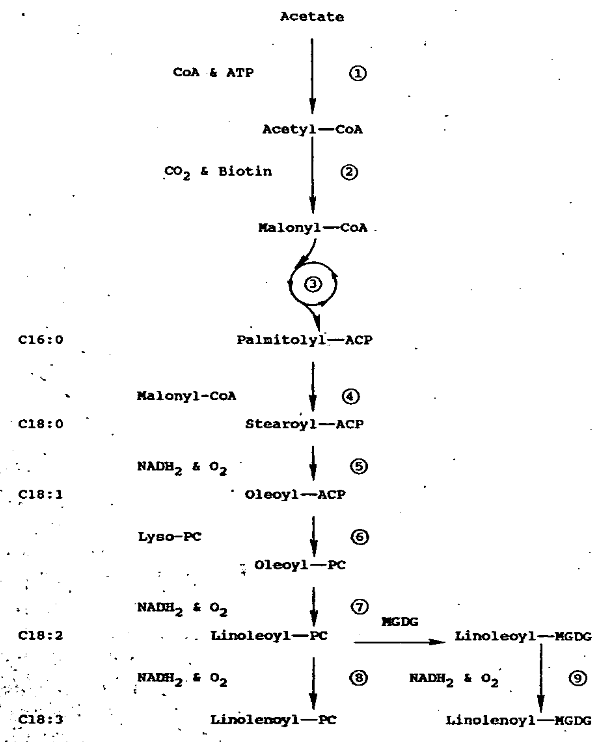


Figure 1. *A scheme showing the biosynthetic pathway of polyunsaturated fatty acids according to the references given in text.* The enzymes catalyzing individual reactions are: (*1*), acetyl-CoA synthetase; (*2*), acetyl-CoA carboxylase; (*3*), fatty acid synthetase; (*4*), fatty acid elongase; (*5*), C18:0 desaturase; (*6*), lysophosphatidylcholinetransferase; (*7*), C18:1 desaturase; (*8*) and (*9*), C18:2 desaturase(s).

can be inhibited by cyanide (Nagai and Bloch, 1968; Jaworski and Stumpf, 1974). In plants, ferredoxin acts as an intermediate electron carrier, transporting two electrons from NADH or water to the double bond being acted on by the desaturase.

Stearoyl-ACP desaturase has been isolated and purified from developing safflower seeds. It was found to be a dimer with a molecular mass of 68 kDa, it required 400 µM oxygen for maximal activity and was stimulated several fold by catalase (McKeon and Stumpf, 1982). In developing seeds as in leaves, the desaturation of stearic acid to oleic acid apparently occurs in plastids (Stumpf, 1980).

The elucidation of the biosynthesis of two polyunsaturated fatty acids, namely linoleate and linolenate, is based mostly on intact tissue investigation or experiments with membrane fractions such as microsomes. Attempts to obtain systems which can be purified and thereby fully characterized have been unsuccessful, but desaturation of oleic to linoleic acids likely occurs outside the chloroplast in both photosynthetic and non-photosynthetic tissues (Roughan and Slack, 1982). Studies have demonstrated that intact chloroplasts only synthesize oleic acid, although small amounts of linoleic and linolenic synthesis have been noted (Heinz *et al.*, 1979). The resulting oleic acid in chloroplasts is transferred to the cytoplasm and esterified to phosphatidylcholine (PC), which acts as the substrate for the desaturation to linoleic acid. The oleoyl-PC desaturase is membrane bound and localized in the endoplasmic reticulum (ER) (Abdelkader *et al.*, 1973; Dubacq *et al.*, 1976; Slack *et al.*, 1976; Trémolières *et al.*, 1980 *a*). It requires NADH and O_2 for activity and it is inhibited by cyanide (Stymne and Appelquist, 1978). In developing seeds, the activity of this enzyme is relatively easily measured in crude cell-free homogenates and microsomal fractions (Stymne and Appelquist, 1980). The desaturation of oleoyl-PC in isolated microsomes from young pea leaves was found to occur predominantly on the *sn*-2 position of PC (Murphy *et al.*, 1985); whereas, the desaturation of oleoyl-PC in isolated potato tuber microsomes was found to occur on both positions although again mostly on the *sn*-2 position (Demandre *et al.*, 1986). However, the isolation and purification of this enzyme has not yet been achieved.

Information concerning the desaturation of linoleate to linolenate is currently very limited. The microsomal desaturation product, linoleoyl PC, is transferred back to the chloroplast. A purified PLTP has been demonstrated as being capable of carrying phospholipids from microsomes to intact chloroplasts (Trémolières *et al.*, 1980 *b*; Drapier *et al.*, 1982; Ohnishi and Yamada, 1982; Dubacq *et al.*, 1984; Grechkin *et al.*, 1984). There has been considerable discussion as to whether the substrates of linoleoyl desaturase are phospholipids, fatty acid CoA forms, or galactolipids (Stumpf, 1980; Roughan and Slack, 1982). In plant leaf tissues, both galactolipids and phospholipids are thought to serve as the substrates (Roughan and Slack, 1984; Williams *et al.*, 1983); while in developing seeds, phosphatidylcholine is believed to be the substrate (Slack *et al.*, 1979). Linoleoyl desaturase resembles other known desaturases in that NADH and O_2 are required for activity and its activity is inhibited by cyanide. This enzyme is considered to reside in chloroplasts in leaf tissues. The activity of a linoleoyl desaturase has been demonstrated to occur in thylakoids of pea chloroplasts (Grechkin *et al.*, 1984). However evidence for more than one site for the formation of 18:3 in plant cells has been accumulating (Frentzen, 1986).

Though plants accumulate large amounts of linolenic acid, the activity of linoleoyl desaturase is very low and less stable relative to other desaturases. Few reports have dealt with the *in vitro* assay of linoleoyl desaturase activity. The activity of linoleoyl desaturase can be assayed by feeding [^{14}C]linoleic acid to intact plants or homogenates. Catalase stimulates the activity of this enzyme (Browse and Slack, 1981). Differential centrifugation of soybean homogenates caused a complete loss of its activity (Stymne and Appelquist, 1980), but the enzyme from linseed cotyledons has been partially stabilized and shown to be located in microsomes (Browse and Slack, 1981). Attempts to further isolate the enzyme have not been successful.

Procaryotic and eucaryotic pathways for the formation of polyunsatured fatty acids

Two different pathways for the formation of lipids in higher plants have been proposed (Roughan *et al.*, 1980; Heinz and Roughan, 1982; Gounaris *et al.*, 1986; Heemskerk *et al.*, 1987). The characteristics of the two pathways are originally based on the positionally specific distributions of the fatty acids between the two positions of the glycerol backbone (Heinz, 1977). In the procaryotic (intraplastidic) pathway, galactolipid synthesis begins with the assembly of predominantly 18 carbon fatty acids on glycerol-3-P forming lysophosphatidic acid followed by incorporation of 16 carbon fatty acids at the *sn*-2 position forming 1-18-2-16-phosphatidic acid (PA) (Sauer and Heise, 1982). The PA is cleaved by a phosphatase, yielding diacylglycerol which is subsequently galactosylated to form MGDG (Frentzen *et al.*, 1982). This group of lipids is exclusively esterified with C16 fatty acids at the *sn*-2 position, while the *sn*-1 position contains C18- and to a lesser extent C16-acyl groups. Since this distribution corresponds to the typical fatty acid pattern of glycerolipids from cyanobacteria, it is called the procaryotic pathway. This pathway is located in chloroplasts and uses galactolipid intermediates as precursors (Roughan and Slack, 1982). The other system is the eucaryotic pathway which forms glycerolipids having C18-fatty acids at both positions. C16-acyl groups are excluded from the *sn*-2 position of lipids formed by the eucaryotic pathway (Williams *et al.*, 1983). This eucaryotic pattern is characteristic of glycerolipids from extraplastidic membranes. Subsequent evidence indicates that the eucaryotic pathway occurs in the cytosol phase and involves microsomal PC as the substrate (Frentzen, 1986).

The basis of the two pathway hypothesis is that fatty acids synthesized *de novo* in the chloroplast may either be used directly for production of chloroplast lipids via the procaryotic pathway (Roughan *et al.*, 1980; Sparace and Mudd, 1982; Heinz and Roughan, 1983), or be exported to enter the eucaryotic pathway at an extrachloroplastic site, particularly in the endoplasmic reticulum (Block *et al.*, 1983; Dubacq *et al.*, 1983; Oursel *et al.*, 1987). The diacylglycerol moiety of PC synthesized by the eucaryotic pathway is returned to the chloroplast probably by the action of a PLTP, where it contributes to the production of thylakoid lipids (Ohnishi and Yamada, 1982; Dubacq *et al.*, 1984). In the eucaryotic pathway, molecular species of PC produced in the microsomes, composed mainly of 18:2 and 18:3 at both the *sn*-1 and *sn*-2 positions, serve as a precursor of MGDG synthesis in the chloroplast (Norman and St. John, 1986).

These two different pathways for the synthesis of plant lipids have been suggested to be associated with the production of polyunsatured fatty acids. This theory is based on the accumulating data from [^{14}C]acetate, $^{14}CO_2$, [^{3}H]glycerol and [^{14}C]oleate labelling of leaves and algae cells *in vivo* (Appleby *et al.*, 1971; Williams and Khan, 1982); from labelling experiments with isolated chloroplasts and microsomal fractions *in vitro* (Roughan *et al.*, 1980; Dubacq *et al.*, 1983); and from enzymological studies (Joyard and Douce, 1977; Douce and Joyard, 1979; Block *et al.*, 1983; Frentzen *et al.*, 1983, 1984). In the procaryotic pathway, 18:1/16:0 monogalactosyl diacylglycerol (MGDG) is synthesized within chloroplasts and desaturated *in situ* to form 18:3/16:3 MGDG (Siebertz *et al.*, 1980). An eucaryotic pathway involving desaturation of microsomal PC provides the diacylglycerol (DAG) precursors for 18:3/18:3 MGDG synthesis (Roughan and Slack, 1984). For example, in *Arabidopsis thaliana*, 18:2/16:2 MGDG (procaryotic pathway) is the substrate for production of 18:3 at the *sn*-1 position of MGDG. The desaturation of 18:2 on PC (eucaryotic pathway) provides 18:2/18:3 PC as a precursor for 18:3/18:3 MGDG synthesis (Norman and St. John, 1986).

According to the positional and fatty acid specificities of the glycerophosphate and monoacylglycerolphosphate acyltransferase, phosphatic acids with a procaryotic pattern are formed in the chloroplast envelope (Stobart *et al.*, 1983; Stymne and Stobard, 1984*a* and *b*, 1985). This phosphatic acid serves as the substrate for the subsequent biosynthesis of monogalactosyldiacylglycerol as well as phosphatidylglycerol. The ability to form procaryotic glycerolipid is decisively controlled by the activity of the plastidial phosphatidic acid phosphatase (Gardiner *et al.*, 1982; Heinz and Roughan, 1983). *In vitro* labelling experiments with isolated chloroplasts from different 16:3 and 18:3 plants indicate that the phosphatase activity is highly correlated with the

amount of 16:3. In the 16:3-plant spinach, this phosphatase is located in the outer membrane of the chloroplast (Joyard and Douce, 1977).

The ER is the primary site of the biosynthesis of eucaryotic glycerolipids in the extraplastidic part of the cell (Moore, 1982; Sauer and Robinson, 1985). The activities of the ER are relevant to the biogenesis of the cells' entire membrane system. However, the microsomal system from photosynthetically active tissues is less well characterized than that from developing seeds. PC has a decisive metabolic role not only in the biosynthesis of eucaryotic diacylglycerolipids in leaves, but also in the biosynthesis of polyunsaturated TAGs in developing seeds (Slack *et al.*, 1978; Roughan and Slack, 1984). The phosphatidic acid formed in the ER membrane serves as a precursor for the biosynthesis of the different phospholipids which can subsequently be used for linoleic and linolenic acid biosynthesis. The desaturation of acyl groups in the ER is conclusively demonstrated to occur on those esterified to PC (Roughan and Slack, 1984). This desaturation occurs in membrane fractions not only from developing oilseeds, but also from photosynthetic tissues.

The relative contribution of these two pathways to MGDG synthesis establishes the fatty acid composition in a given plant species (Roughan, 1985). For example, if only the eucaryotic pathway is used, the plant is classified as an "18:3" species. If the procaryotic pathway contributes substantially to cellular lipids, the plant is termed a "16:3" species. A wide variation in the percentage of 16:3 in MGDG has been observed among plant species (Jamieson and Reid, 1971) and this diversity is considered to reflect the relative activity of enzymes involved in the two separate biosynthetic pathways (Roughan, 1975).

The glycerolipid biosynthesis capacity in the two pathways must be well balanced to result in the different levels of procaryotic and eucaryotic glycerolipids in the membrane system of chloroplasts of various plants (Douce and Joyard, 1979). In *A. thaliana*, almost equal amounts of chloroplast lipids are synthesized by the procaryotic and eucaryotic pathways (Browse *et al.*, 1986*b*). The quantities of individual lipids produced by the two routes are very different. Chloroplast phosphatidylglycerol is synthesized via the procaryotic pathway, whereas chloroplast PC is a product of the eucaryotic pathway. In one *A. thaliana* mutant (JB1), the synthesis of 18:3 from the procaryotic pathway is deficient, but plants compensate by producing more 18:3 from the eucaryotic pathway (Norman and St. John. 1986).

The production of glycerolipids in the two biosynthetic pathways is also modulated by the concentrations of glycerol 3-phosphate (Gardiner *et al.*, 1982). This regulation probably results from the differing affinities of the glycerol 3-phosphate acyltransferase for the acyl acceptor. Results of *in vivo* and *in vitro* labelling experiments showed that an increased cellular concentration of glycerol 3-phosphate in leaves of 16:3 and 18:3-plants had no effect on the total incorporation of acetate into lipids, but significantly stimulated the synthesis of procaryotic glycerolipids (Gardiner *et al.*, 1982).

Synthesis of polyunsaturated fatty acids in seed triacylglycerols

The fatty acid composition of TAGs in oilseeds is species and often variety specific (Hilditch and Williams, 1964; Downey and McGregor, 1975). The relative proportions of the constituent fatty acids esterified at the three positions of the glycerol molecule also differ considerably. In general, the unsaturated C18 fatty acids, oleic, linoleic, and linolenic acids are major constituents of the TAGs of edible oilseed crops.

PC plays an important metabolic role during the formation of polyunsaturated triacylglycerols in seeds (Wilson *et al.*, 1980). In developing cotyledons, labelled fatty acids accumulate rapidly into PC and diacylglycerols, but initially only at a slow rate into TAG. During a chase, following pulse-labelling, radioactivity is lost largely from the oleic acid of this phospholipid and accumulates in the polyunsatured C18 fatty acids, linoleate and linolenate of triacylglycerols (Dybing and Craig, 1970; Slack *et al.*, 1978). *In vitro* the microsomal desaturase from developing cotyledons uses oleoyl and linoleoyl PC as substrates to form linoleoyl and linolenoyl PC, respectively, as products (Stymne and Appelquist, 1978; Slack *et al.*, 1979; Browse and Slack, 1981). Consequently, this phospholipid appears to serve as a donor of these fatty acid for TAG formation (*fig.* 2). Other phospholipids, in addition to PC, can also serve as acyl donors (Wilson *et al.*, 1980). Since labelled glycerol moieties as well as acyl moieties were transferred from PC to TAG during a chase, it is suggested that this phospholipid could provide both the DAG and the fatty acids from which TAG is formed (Slack *et al.*, 1978).

The microsomal fraction from oilseeds possesses all the necessary enzyme activities for *de novo* biosynthesis of TAGs, namely glycerol 3-phosphate acyltransferase, monoacylglycerol 3-phosphate acyltransferase, phosphatidic acid phosphatase and diacylglycerol acyltransferase (*fig.* 2). PC formed in the microsome is used as a substrate for the subsequent desaturation of the oleoyl groups esterified at the *sn*-1 as well as the *sn*-2 position of the glycerol

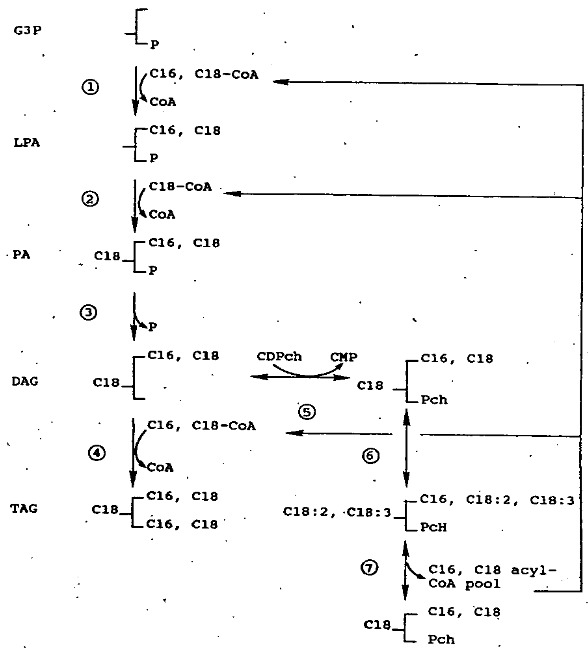


Figure 2. *A scheme showing the biosynthetic pathway of polyunsaturated triacylglycerols in developing oilseeds according to the references given in text.* C16 is mostly C16:0 whereas C18 can be C18:0, C18:1, C18:2, or C18:3. The enzymes catalyzing individual reactions are: (*1*), G3P acyltransferase; (*2*), 1-acylglycerol 3-P acyltransferase; (*3*), PA phosphatase; (*4*), DAG acyltransferase; (*5*), DAG choline phosphotransferase; (*6*) C18:1 and C18:2 PC desaturases and (*7*), lyso-PC acyltransferase.

backbone (Slack *et al.*, 1979; Rochester and Bishop, 1984; Stobart and Stymne, 1985). The lysophosphatidylcholine acyltransferase in the microsomal fraction exclusively attacks the *sn*-2 position of PC. It possesses a high specificity for unsaturated C18-fatty acids and a slight preference for C18:1 (Stobart *et al.*, 1983; Stymne *et al.*, 1983). Hence this enzyme preserves the eucaryotic fatty acid pattern but affects an acyl exchange between acyl-CoA and position 2 of PC by the combined backward and forward reactions. The biosynthetic pathway making polyunsaturated fatty acids of TAGs via PC can be channeled to TAGs by the reverse reaction of cholinephosphotransferase. These combined forward and backward reactions alter the acyl-CoA mixture exported from the plastids, resulting in a decrease of C18:1 and an increase of C18:2 or C18:3 groups corresponding to the fatty acid sensitivity of the acyltransferase and the fatty acid composition at position 2 of PC, respectively. The acyl exchange coupled to the DAG PC equilibrium gives rise to a continued enrichment of the glycerol backbone with polyunsaturated fatty acids (Griffiths *et al.*, 1985; Stobart and Stymne, 1985; Stymne and Stobart, 1985). By this pathway the whole DAG moiety of PC is incorporated into TAGs (Stymne and Stobart, 1984 *b*).

In maturing soybean seeds, the formation of linolenate is developmentally regulated. The amount of linolenic acid is highest during the very early stages of seed formation with the relative amount decreasing at the later stages of development (Reubel *et al.*, 1972; Roehm and Privette, 1970; Cherry *et al.*, 1984). Assays of cell-free extracts have demonstrated that the homogenates of early stage cotyledons possess higher and more stable linoleoyl desaturase activity than those of later stages (Stymne and Appelquist, 1980).

Manipulation of the synthesis of polyunsaturated fatty acids

Genetic alteration of the synthesis of polyunsaturated fatty acids

Substantial variation occurs among species in the level of polyunsaturated fatty acids in seed oil. Some species, such as sunflower (*Helianthus annus*) and safflower (*Carthamus tinctorius*) contain essentially no linolenic acid, but have high levels of linoleic acid. However others, including soybean (*Glycine max*) rapeseed/canola (*Brassica napus* and *B. campestris*) and flax (*Linum usitatissimum*), all contain significant quantities of linolenic acid (Downey, 1987). The content and composition of polyunsaturated fatty acids in lipids is genetically regulated in plants. The genetic control of polyunsaturated fatty acid synthesis is best studied by isolation and characterization of mutants with altered formation of these fatty acids. Such mutants have been isolated from various plant species by means of physical and chemical mutagenesis. Among these are mutants from *A. thaliana* (Browse *et al.*, 1986 *a*), flax (Green and Marshall, 1984; Green, 1986), soybeans (Wilson *et al.*, 1981; Hammond and Fehr, 1983; Wilcox, *et al.*, 1984), and *Brassica* oilseed crops (Rakow, 1973; Robbelen and Nitsch, 1975) which have been studied in some detail.

The *A. thaliana* fatty acid desaturation mutants isolated and characterized by Browse, Somerville and coworkers (Browse *et al.*, 1984, 1985, 1986 *a*; Somerville *et al.*, 1987) have particular promise in facilitating the elucidation of the molecular genetic controls and functional significance of unsaturated fatty acids in plant leaves. Four mutants (designated *fadA*, *fadB*, *fadC* and *fadD*) isolated by direct analysis of fatty acid composition of leaf tissues from

an ethylmethane sulfonate (EMS)-mutagenized population of plants (Browse *et al.*, 1986 *a*) have been tentatively described in terms of the sites of the particular enzymatic lesions (Somerville *et al.*, 1987). The *fadA* mutants apparently lack the desaturase which converts 16:0 to *trans*-3-hexadecenoic acid. The *fadB* mutants accumulate high levels of 16:0 but are deficient in 16:1, 16:2 and 16:3 fatty acids. The wildtype and the mutants *fadC* and *fadD* contain almost identical C16/C18 ratios. However, the fatty acid composition of *fadD* shows a decreased level of both C18:3 and C16:3 with a corresponding increase of C18:2 and C16:2 in comparison to the wildtype. The mutant *fadC* also contains reduced levels of polyenoic acids, but in this mutant it is the monoenoic acids which show a corresponding increase. These results imply that the desaturase activities affected by the mutations in *fadC* and *fadD* can act on both C16 and C18 acyl groups, and the site for the insertion of a new double bond is determined from the methyl end of the chain. In chloroplasts, the biosynthesis of dienoic and trienoic acids is therefore catalyzed by *n*-6 and *n*-3 desaturase activities (Frentzen, 1986).

The kinetics of in vivo labelling of lipids with [^{14}C]acetate and quantitative analysis of the fatty acid composition of individual lipids suggest that reduced activity of a glycerolipid *n*-3 desaturase is responsible for the altered lipid composition of the *fadD* mutant (Browse *et al.*, 1986 *a* and *b*). The effects of the mutation are fully expressed when plants are grown at temperatures above 26°C, but are relatively minor below 18°C, suggesting a temperature sensitive enzyme. Both chloroplast (16:3 containing) and extrachloroplast (18:3 in extrachloroplast membranes) lipids are equally affected by the mutation, indicating that either the desaturase is located both outside and inside the chloroplast or C18:3 formed inside the chloroplast is re-exported to other cellular sites.

Studies on the synthesis of unsaturated MGDG molecular species of the *fadD* mutant suggest that multiple substrates are involved in the desaturation of linoleic acid to linolenic acid for the production of unsaturated galactolipids (Norman and St. John, 1986). The mutation selectively reduces the levels of 18:3/16:3 and increases the amount of 18:3/18:3 despite the overall reduction in 18:3, suggesting that a chloroplastic pathway for desaturation at the *sn*-1 position of MGDG utilizes 18:2/16:2 MGDG as the substrate. This procaryotic pathway is apparently deficient in this mutant. The eucaryotic pathway desaturating 18:2 to 18:3 at the *sn*-2 position of PC predominates in the mutant. Genetic characterization of the *fadD* mutation showed that the low trienoic fatty acid content is controlled by a single recessive nuclear gene. There is no change in the fatty acid composition of seed and root lipids in this mutant (Browse *et al.*, 1986 *a*), indicating that different pathways or isozymes may operate in the different tissues.

The nutritional and industrial value of seed storage lipids is dependent primarily upon the fatty acid composition. Of particular importance is the relative proportion of the C18 unsaturated fatty acids namely oleic, linoleic and linoleic acids (Smouse, 1979). Oils with high levels of polyunsaturated fatty acids particularly linolenic acid are less suitable for use as cooking oils due to the poor oxidative stability of these fatty acids. Efforts have therefore been expended on altering the fatty acid composition of seed storage lipids to meet the desired end use of the oils [it should be noted however, that 18:3 (an omega-3 fatty acid) has recently been implicated as playing an important role in human health (Booyens and van der Merwe, 1985)].

Searching for and using genetic variants of fatty acid composition have become common approaches to achieving the manipulation of fatty acid composition in seed oil crops. For example, flax oil contains a high percentage of 18:3 fatty acid (45%-65%). This high level precludes its use as an edible oil and gives its traditional industrial use. Genotypes with 2% 18:3 have been isolated (Green, 1986). This alteration is achieved by selection within the F2 generation of a cross between two induced mutants with reduced levels of linolenic acid (28%-30%) (Green and Marshall, 1984). This near elimination of linolenic acid from the seed lipids is accompanied by a comparative increase in the content of linoleate, with the proportions of other fatty acids remaining unaltered. These results indicate that the mutations block the final desaturation of linoleic to linolenic acid. Genetic analysis of crosses among these mutants and their parental cultivar revealed that these mutations are in different unlinked genes and exhibit additive (codominant) gene action. Two genetic loci with additive effects have therefore been identified to control the linolenate content (Green, 1986).

Soybean seed oil is the most common edible oil in the world (Smith, 1981). About eight percent is composed of 18:3 and this relatively high content of linolenic acid as well as linoleic acid has been considered to be an important factor lowering the autooxidative and flavor stability of soybean oil. Accessions of the commercial soybean species (*Glycine max.*) display 18:3 content of 4-15% of the seed oil. In other species of the genus *Glycine*, the linolenate content ranged from 11.3-27.2% (Smith, 1981; Chaven *et al.*, 1982). The lack of genotypes with very low 18:3 content within the genus *Glycine* has spurred other approaches to reduce the 18:3 content in soybean seed oil.

Thus far, the most effective approach to develop soybean strains with genetically determined low levels of linolenic acid has been the use of chemical mutagens. Treatment with EMS significantly increased the variation in the fatty acid composition in the soybean cultivar "Century". A genetically stable mutant designated C1640 with 3.4% 18:3 was identified (Wilcox *et al.*, 1984). This mutation is controlled by a single gene locus (Wilcox and Cavins, 1985) designated *Fan* (Wilcox and Cavins, 1987). Treatment of one low linolenic acid breeding line after recurrent selection with EMS resulted in a line, designated A5, with linolenic acid content of 2.9-4.1% (Hammond and Fehr, 1983, 1984). Graef *et al.* (1988) demonstrated that fatty acid composition should be considered a quantitative character in crosses which involve A5 as a parent. The decreased 18:3 content in A5 is apparently the result of a decreased rate of 18:1 desaturation of oleoyl-PC in this genotype. Interestingly we find that both C1640 and A5 have reduced 18:3 in root lipids but not in any other vegetative tissues (Wang *et al.*, 1988). A5 had a corresponding increase in 18:2 in the root lipids in contrast to the increase in 18:1 in the seed lipids (Wang *et al.*, 1988).

Recurrent selection for high oleic acid and a high ratio of 18:1 to 18:2 + 18:3 has been conducted to alter the fatty acid composition of soybean oil (Wilson *et al.*, 1981). The results of this study showed that the 18:1 content of the oil increased linearly from 24.8 to 33.0%. The inversely correlated trait, 18:3, was reduced linearly from 7.8 to 6.3%.

An objective of rapeseed (8-10% 18:3) breeding has been to reduce the level of linolenic acid to less than 3% while maintaining or increasing the level of the nutritionally desirable linoleic acid presently at 20-25%. [It should be noted, however, that the major objective of the genetic improvement of rapeseed oil has been to reduce the erucic acid content (22:1). Breeding efforts toward these goals have been highly successful resulting in the development, for instance, of the new Canola type rapeseed oil with a reduction of erucic acid from about 50% to less than 1% and a corresponding increase in oleic acid (18:1) (Downey, 1987). The canola oils are acceptable for many edible purposes even though they have an 18:3 content averaging 10% because of the high 18:1/18:2 + 18:3 ratio.] Chemical mutagenesis was successful in reducing the linolenic acid level to 5.5% (Rakow, 1973) and subsequent selection produced material with a linolenate level as low as 3.2% (Robbelen and Nitch, 1975). In the *Brassica* species, linolenic acid biosynthesis was observed to occur only in those seeds which possess green photosynthetically active chloroplasts during certain stages of their development (Thies, 1970).

Studies of the metabolism of MGDG molecular species in *A. thaliana* leaf mutants have revealed different sites and substrates for linolenic acid synthesis (Norman and St. John, 1986). The use of other mutants with low seed 18:3 for elucidating the mechanism of regulation and synthesis of linolenate is however still at preliminary stages. Despite great efforts having been expended on the selection of low 18:3 mutants in soybeans, no 18:3-null mutants have been found. It is not yet clear if a certain amount of 18:3 is essential for soybean seed development. However, the near absence of 18:3 in other oilseed crops diminishes this possibility. Inconsistent expression of the low 18:3 trait in various tissues has been observed in soybean seed mutants (Martin and Rinne, 1985). The reduction of linolenate content in some mutants is due to the blockage of desaturation at 18:1, whereas others occur at the 18:2 desaturation step (Cherry *et al.*, 1984). Nevertheless, these mutants have provided an instrumental approach to the study of 18:3 biosynthesis and regulation.

Chemical modulation of the formation of polyunsaturared fatty acids

One substituted pyridazinone, San 9785 (4-chloro-5-dimethylamino-2-phenyl-3(2H) pyridazinone), is a potent inhibitor of the desaturation of linoleate to linolenate (St. John, 1976; Lem and Williams, 1981). San 9785 has been shown to selectively affect the levels of linolenate in several species of higher plants without causing any gross change in leaf development and chloroplast content (Laskay *et al.*, 1983). Labelling studies with fatty acid precursors, such as $^{14}CO_2$ and [^{14}C]acetate, have demonstrated a reduction in linolenate radiolabelling in the presence of San 9785 and a concomittant increase in the levels of [^{14}C]linoleate (Willemot *et al.*, 1982). Therefore, San 9785 has been considered to have a direct effect upon the conversion of linoleate to linolenate. This compound has been widely used in the manipulation and study of the synthesis and function of linolenic acid.

It is suggested that San 9785 inhibits 18:3 formation at the procaryotic pathway with little effect on the eucaryotic one (Lem and Williams, 1981; Norman and St. John, 1987). San 9785 was shown to reduce linoleate desaturation of MGDG but not PC. The differential effects of San 9785 on the pathway of MGDG synthesis was studied in *A. thaliana* (Norman and St. John, 1987). 18:3/16:3 MGDG was decreased by San 9785, and 18:2/16:3 and 18:2/16:2 MGDG concurrently increased. Kinetic studies using exogenously incorporated [^{14}C]18:1 indicated that

18:3/18:3 MGDG originated from a 18:2/18:3 diacylglycerol precursor derived from PC. The formation of 18:3 at the *sn*-2 position of PC was less sensitive to San 9785 than desaturation of 18:2 at the *sn*-1 position of 18:2/18:3 MGDG which is proposed to occur within the chloroplasts.

Direct evidence has shown that the site of action of San 9785 on fatty acid biosynthesis in higher plants is at the level of linoleic acid desaturation, but there are large variations in sensitivity between plant species (Hilton *et al.*, 1971; Murphy *et al.*, 1980, 1985). Effects of San 9785 have also been reported upon photosynthetic oxygen evolution (Khan *et al.*, 1979 *a;* Lem and Williams, 1981), thylakoid ultrastructure and chlorophyll-proteins (Khan *et al.*, 1979 *b;* Davies and Harwood, 1983; Leech and Walton, 1983). It has been reported that, whereas San 9785 inhibits the incorporation of [^{14}C]acetate into linolenate in spinach leaf discs, it had no effect upon this incorporation in either isolated chloroplasts or whole leaves of spinach (Willemot *et al.*, 1982). The effects of San 9785 upon photosynthetic oxygen evolution that were reported from both *Vicia faba* leaf discs and isolated spinach chloroplasts were not found in the case of whole leaves of barley. It therefore appears that San 9785 may be quite variable in its effects upon plants depending on the species studied and the pretreatment of the tissues.

Uptake studies demonstrated that the uptake of San 9785 was a reflection of water uptake (Murphy *et al.*, 1985). Following its uptake, San 9785 was rapidly converted into other compounds in pea, but only gradually metabolized in cucumber and ryegrass. The differential sensitivity of higher plants to San 9785 was shown to be due to variation both in uptake and in metabolism.

San 9785 reduces the 18:3 content in soybean cotyledons developing *in vitro* (St. John *et al.*, 1984; Wang *et al.*, 1987 *a*). It also decreases the activity of lipoxygenase, an enzyme catalyzing the oxidation of polyunsaturated fatty acids, in peanut and soybean seeds (Ory *et al.*, 1981, 1984; St. John *et al.*, 1984; Wang and Hildebrand, 1987). Treatment with San 9785 causes these changes without affecting the yield and other important agronomic parameters. Thus, it is suggested that this compound could be applied to the improvement of soybean quality.

The mode of action of San 9785 on the inhibition of linoleate desaturation is not currently understood. The observation that San 9785 had little effect on linolenate synthesis in isolated chloroplasts (Willemot *et al.*, 1982) suggests that either protein synthesis is required for San 9785 action or it needs to be metabolized first in the cytosol in order to be functional. In addition, the decrease of radioactive labelling of linolenate in treated tissues is often seen after prolonged incubation with San 9785 (Lem and Williams, 1981; Davies and Harwood, 1983). This delay may be due do the slow uptake of San 9785; to a delay in the conversion to active metabolites (St. John and Hilton, 1976); to an inhibition of the synthesis of linoleate desaturase; and/or to an elevation of degradation of the desaturase by this compound. Furthermore, the inhibition of linoleate desaturation by San 9785 does not occur in the presence of cycloheximide (Norman, personal communication), a cytosol protein synthesis inhibitor, indicating that protein synthesis is required for San 9785 function.

Environmental effects on the synthesis of polyunsaturated fatty acids

The synthesis of linolenic acid in plants can be affected by a number of environmental factors which include temperature, light, water stress and salt stress (Harwood, 1984; Trémolières, 1985). Low temperature often stimulates the synthesis of polyunsaturated fatty acids in various plant tissues (Hazel and Prosser, 1974). Plants grown at low temperatures during seed maturation accumulate more 18:3 in TAGs than those grown at high temperatures (Reubel *et al.*, 1972; Hawkins *et al.*, 1983 *a* and *b;* Cherry *et al.*, 1985). A change in the ambient temperature caused a marked alteration over a 24 h period in the proportions of unsaturated C18 fatty acids in PC and DAG during soybean and linseed cotyledon development (Slack and Roughan, 1978). At high temperatures, 18:1 increased and 18:2 and 18:3 decreased. For soybean cultivars with different levels of linolenate grown in Northern areas (low temperature) and Southern areas (high temperature), seeds produced in the North are significantly higher in myristate and linolenate, but are lower in oleate (Cherry *et al.*, 1985). Trémolières *et al.* (1978, 1982) found that in rapeseed low temperatures increased the level of polyunsaturated fatty acids at the expense of oleic acid biosynthesis without change of the total lipid content. Similar effects of temperature on polyunsaturated fatty acid synthesis have also been observed in plant cells in culture (Trémolières *et al.*, 1978; MacCarthy and Stumpf, 1980; Trémolières *et al.*, 1982). However, in developing sunflower seeds (Trémolières *et al.*, 1982) low temperatures decreased lipid accumulation with little change in fatty acid composition. Linolenic acid biosynthesis in *Pharbitis nil* cotyledons was likewise very slow at lower temperatures (17C) compared to higher temperature (27C).

There is some controversy concerning the reason for the elevated lipid desaturation in plants grown

at low temperatures. Three schools of thought exist. Some believe that low growth temperature results in an alteration of the activity or amounts of the desaturase enzymes themselves and that such changes are adaptive, enabling cellular membranes to function more effectively (Thompson, 1980). Others propose that the low growth temperature increases oxygen solubility, therefore providing more substrate for the existing desaturase enzymes (Rebeillé *et al.*, 1980). Still others (Browse and Slack, 1983) found evidence which suggests that the apparent increased lipid desaturation at lower temperature in developing safflower cotyledons is actually the consequence of greater increases in fatty acid synthesis than oleate desaturation at higher temperatures which therefore decreases the ratio of polyunsaturated to monounsaturated fatty acids at higher temperatures.

Light also has a profound impact on linolenic acid synthesis. Linolenic acid, esterified to phospholipids or galactolipids, is the principal component of the lipid matrix of photosynthetic membranes of chloroplasts (James and Nichols, 1966), and it can account for up to 90% of the total fatty acids in that organelle (Leech and Murphy, 1976; Trémolières, 1985). The easiest and most widely used experimental system is the greening of etiolated tissues. Etioplasts contain much lower levels of linolenate than chloroplasts (Trémolières and Lepage, 1971; Nichols, 1965; Trémolières and Mazliak, 1970). Darkgrown pea seedlings are rich in linoleic acid. After illumination of these seedlings, a very significant increase in linolenic acid is observed in the young leaf sections, whereas only small variations are seen in fatty acid composition of other sections (Trémolières and Lepage, 1971). Studies have also shown that photoautotrophic cells in culture produce much more linolenic acid than heterotrophic cells (Husemann *et al.*, 1980).

It was found that greening cucumber cotyledons exhibited a dramatic increase in the ability to desaturate exogenously added [^{14}C]linoleic acid (Murphy and Stumpf, 1979). The inhibition of the light-dependent increase in desaturating activity by cycloheximide suggests that this process is dependent on protein synthesis on the 80S ribosomes (*i.e.* cytoplasmic), which parallels similar findings in other light-induced systems. However, oleate and linoleate desaturation in leaves of maize seedlings was largely independent of the previous light treatment of the seedlings (Hawke and Stumpf, 1980); there was no evidence of light-induced desaturase activities. These results are in sharp contrast to those observed with developing cucumber cotyledons. *In vivo* desaturase activity was present in tissues of widely different levels of differentiation and chlorophyll content obtained from light grown maize seedlings.

Water stress on plants likewise results in a change of linolenic acid synthesis (Pham-Thi *et al.*, 1982). Drought appears to reduce the ability of plants to synthesize 18:3. Studies have shown a decrease in this polyunsaturated fatty acid in cotton leaves submitted to water stress by withholding irrigation. Experiments on incorporation with [^{14}C]acetate as the precursor clearly indicate that water deficits provoke a severe inhibition of unsaturated fatty acid biosynthesis. The inhibition of oleate and linoleate desaturation by drought certainly contribute to the decrease in the content of leaf polyunsaturated fatty acids observed in water-stressed cotton leaves (Ouedraogo *et al.*, 1984; Pham-Thi *et al.*, 1985, 1987).

The effect of salt stress such as sodium chloride on lipids is expressed mainly by a decrease of the linolenic acid content (Zarrouk and Cherif, 1984). It seems that a variety of environmental factors can affect polyunsaturated fatty acid synthesis, either directly or indirectly. However, the molecular mechanisms of such changes are obscure, and it is not yet known what role these changes in lipids may play in the adaptation of plants to such environmental stresses.

Conclusion

This review on the current understanding of linolenate production indicates that the process of the biosynthesis of linolenic acid is complex and many fundamental questions remain to be answered. In particular, further studies are needed to establish the substrates for 18:2 desaturation, existence of multiple distinct desaturases, subcellular sites of the desaturation, enzymatic and molecular genetic regulation, and the coordination of different pathways for linolenic acid formation.

Regarding the manipulation of linolenate content in oilseed crops, conventional oilseed breeding has been primarily concerned with the genetic expression of fatty acid composition. But the more that is known of fatty acid and lipid synthesis, the more effectively the breeder can assess the opportunities for oil quality improvement and design the appropriate breeding strategies. Moreover, the rapid development of biotechnology opens additional avenues for the genetic engineering of fatty acid composition of oilseed crops (Knauf, 1987). The scope of possibilities of manipulating storage fatty acids and lipids using genetic engineering is indicated by the diversity of different fatty acid compositions which exist in various types of oilseeds. Plant breeders have already shown that the composition of polyunsaturated fatty acids of two cultivars of an

oilseed crop can differ considerably without obvious effects on agronomic suitability. The modification of seed storage fatty acid composition appears to be feasible, but applied research and product development will depend on gaining a better understanding of the biochemical and genetic processes of lipid biosynthesis. Requisite research for the effective application of biotechnological approaches to the manipulation of lipid composition will be the identification and characterization of the key enzymes controlling lipid composition.

In soybeans, one goal of lipid alteration is to decrease the linolenic acid content in order to increase the stability of soybean oil for cooking purposes. The complexity of linolenic acid biosynthesis within plant tissues and within different parts of a cell present difficult challenges to current biotechniques. In addition, a prerequisite in a genetic engineering project is the ability to monitor the gene of interest and its product. Therefore, by using various approaches, studies in several laboratories are currently endeavoring to identify the gene products involved in the control of 18:3 content (Somerville *et al.*, 1987; Wang *et al.*, 1987 *a* and *b*). Further research will be aimed at isolating those genes. The fact that a number of studies have shown that single gene changes can have a large impact on 18:3 content (Wilcox and Cavins, 1985; Green, 1986; Somerville *et al.*, 1987) is encouraging to the prospects of genetic engineering of plants for altered lipid composition. However, progress along these lines will be complicated unless effective protocols are developed for the purification of the key enzymes (such as linoleate desaturases) controlling linolenate levels in plant tissues.

References

Appleby R. S., Safford, R. and Nichols, B., 1971. — The involvement of lecithin and monogalactosyl diglyceride in linoleate synthesis by green and blue-green algae. *Biochim. Biophys. Acta*, **248**, 205-211.

Ben Abdelkader A., Cherif A., Demandre C. and Mazliak P., 1973. — The oleyl-CoA desaturase of potato tubers., *Eur. J. Biochem.*, **32**, 155-165.

Block M. A., Dorne A. J., Joyard, J., and Douce, R., 1983. — The acyl-CoA synthetase and acyl-CoA thioesterase are located on the outer and inner membrane of the chloroplast envelope respectively. *FEBS. Lett.*, **153**, 377-381.

Booyens J. and van der Merwe C., 1985. — Chronic arachidonic acid eicosanoid in balance: A common feature in coronary artery disease, hypocholesterolemia, cancer and other important diseases. *Medical Hypotheses*, **18**, 53-60.

Browse J., McCourt P. and Somerville C. R., 1984. — Glycerolipid metabolism in leaves: new information from *Arabidopsis* mutants. In *Structure, function and metabolism of plant lipids*, Siegenthaler P. A. and Eichenberger W., ed., Elsvier Science Publishers, Amsterdam, 167-170.

Browse J., McCourt P. and Somerville C., 1985. — A mutant of *Arabidopsis* lacking a chloroplast-specific lipid. *Science*, **227**, 763-765.

Browse J., McCourt P. and Somerville C., 1986 *a*. — A mutant of *Arabidopsis* deficient in C18:3 and C16:3 leaf lipids. *Plant Physiol.*, **81**, 859-864.

Browse J. and Slack C. R., 1981. — Catalase stimulates linoleate desaturase activity in microsomes from developing linseed cotyledons. *FEBS Lett.*, **131**, 111-114.

Browse J. and Slack C. R., 1983. — The effects of temperature and oxygen on the rates of fatty acid synthesis and oleate desaturation in safflower (*Carthamus tinctorius*) seed. *Biochim. Biophys. Acta*, **753**, 145-152.

Browse J., Warwick N., Somerville C. R. and Slack C. R., 1986 *b*. — Fluxes through the procaryotic and eucaryotic pathways of lipid synthesis in the "16:3" plant *Arabidopsis thaliana*. *Biochem. J.*, **235**, 25-31.

Carver B. F. and Wilson R. F., 1984. — Triacylglycerol metabolism in soybean with genetically altered fatty acid composition. *Crop Sci.*, **24**, 1020-1023.

Chaven C., Hymowitz T. and Newell C. A., 1982. — Chromosomal number, oil and fatty acid content of species in the genus *Glycine*. *J. Am. Oil Chem. Soc.*, **59**, 23-25.

Chérif A., Dubacq J. P., Mache R., Oursel A, and Trémolières A., 1975. — Biosynthesis of α-linolenic acid by desaturation of oleic and linoleic acids in several organs of higher and lower plants and algae. *Phytochemistry*, **14**, 703-706.

Cherry J. H., Bishop L., Hasegawa P. M. and Leffler H. R., 1985. — Differences in the fatty acid composition of soybean seed produced in Northern and Southern areas of the U.S.A. *Phytochemistry*, **24**, 237-241.

Cherry J. H., Bishop L., Leopold N., Pikaard C. and Hasegawa P. M., 1984. — Patterns of fatty acid deposition during development of soybean seeds. *Phytochemistry*, **23**, 2183-2186.

Davies A. O. and Harwood J. L., 1983. — Effect of substituted pyridazinones on chloroplast structures and lipid metabolism in greening barley leaves. *J. Exp. Bot.*, **34**, 1089-1091.

De la Roche A. I., Andrews C. J., Pomeroy M. K., Weinbager P. and Kates M., 1972. — Lipid changes in winter wheat seedlings (*Triticum aestivum*) at temperature induced cold hardness. *Can. J. Bot.*, **50**, 2401-2405.

Demandre C., Trémolières A., Justin A. M. and Mazliak P., 1986. — Oleate desaturation in six phosphatidyl-

choline molecular species from potato tubers. *Biochim. Biophys. Acta*, **877**, 380-386.

Douce R. and Joyard J., 1979. — Structure and function of the plastid envelope. *Adv. Bot. Res.*, **7**, 1-116.

Downey R. K., 1987. — Genetic manipulation of oilseed quality. In *The metabolism, structure and function of plant lipids*, Stumpf P. K., Mudd J. B. and Nes W. D., ed., Plenum Press, New York, 669-676.

Downey R. K. and McGregor D. I., 1975. — Breeding for modified fatty acid composition. *Curr. Adv. Plant Sci.*, **6**, 151-167.

Drapier D., Dubacq J.-P., Trémolières A. and Mazliak P., 1982. — Cooperative pathway for lipid biosynthesis in young pea leaves: oleate exportation from chloroplasts and subsequent integration into complex lipids of added microsomes. *Plant Cell Physiol.*, **23**, 125-135.

Dubacq J.-P., Drapier D., Trémolières A. and Kader J. C., 1984. — Role of phospholipid transfer protein in the exchange of phospholipids between microsomes and chloroplats. *Plant Cell Physiol.*, **25**, 1197-1204.

Dubacq J.-P., Drapier D. and Trémolières A., 1983. — Polyunsaturated fatty acid synthesis by a mixture of chloroplasts and microsomes from spinach leaves: evidence for two distinct pathways of the biosynthesis of trienoic acids. *Plant Cell Physiol.*, **24**, 1-9.

Dubacq J. P., Mazliak P. and Trémolières A., 1976. — Subcellular localization of the oleoyl-CoA desaturase activity in pea leaves. *FEBS Lett.*, **66**, 183-186.

Dybing C. B. and Craig B. M., 1970. — Fatty acid biosynthesis and incorporation into lipid classes in seeds and seed tissues of flax. *Lipids*, **5**, 422-429.

Frentzen M., 1986. — Biosynthesis and desaturation of the different diacylglycerol moieties in higher plants. *J. Plant Physiol.*, **124**, 193-209.

Frentzen M., Hares M. and Schiburr A., 1984. — Properties of the microsomal glycerol 3-P and monoacylglycerol 3-P acyltransferase from leaves. In *Structure, function and metabolism of plant lipids*, Siegenthaler P. A. and Eichenberge, W., ed., Elsevier Science Publishers, Amsterdam, 105-110.

Frentzen M., Heinz E., McKeon T. and Stumpf P. K., 1982. — *De novo* biosynthesis of galactolipid molecular species by reconstituted enzyme systems from chloroplasts. In *Biochemistry and metabolism of plant lipids*. Wintermans J. F. G. M. and Kuiper P. J. C., ed., Elsevier Biomedical Press, Amsterdam, 141-152.

Frentzen M., Heinz E., McKeon T. and Stumpf P. K., 1983. — Specificities and selectivities of glycerol-3-phosphate acyltransferase and monoacylglycerol-3-phosphate acyltransferase from pea and spinach chloroplasts. *Eur. J. Biochem.*, **129**, 629-636.

Gardiner S. E., Roughan P. G. and Slack R. C., 1982. — Manipulating the incorporation of [1-14C]-acetate into different leaf glycerolipids in several plant species. *Plant Physiol.*, **70**, 1316-1320.

Gounaris K. and Barber J., 1983. — Monogalactosyldiacylglycerol: the most abundant polar lipid in nature. *Trends Biochem. Sci.*, **8**, 378-381.

Gounaris K., Barber J. and Harwood J. L., 1986. — The thylakoid membranes of higher plant chloroplasts. *Biochem. J.*, **237**, 313-326.

Graef G. L., Fehr W. R., Miller L. A., Hammond E. G. and Cianzio S. R., 1988. — Inheritance of fatty acid composition in a soybean mutant with low linolenic acid. *Crop Sci.*, **28**, 55-58.

Grechkin A. N., Gafavora T. E. and Tarchevsky J. A., 1984. — Linoleate Δ15-desaturase activity of pea leaf chloroplasts is localized in thylakoids. In *Structure, function and metabolism of plant lipids*. Siegenthaler P. A. and Eichenberger W., ed., Elsevier Science Publishers, Amsterdam, 51-54.

Green A. G., 1986. — Genetic control of polyunsaturated fatty acid biosynthesis in flax (*Linum usitatissimum*) seed oil. *Theor. Appl. Genet.*, **72**, 654-661.

Green A. G. and Marshall D. K., 1984. — Isolation of induced mutants in linseed (*Linum usitatissimum*) having reduced linolenic acid content. *Euphytica*, **33**, 321-328.

Griffiths G. A., Stobart K. and Stymne S., 1985. — The acylation of *sn*-glycerol 3-phosphate and the metabolism of phosphotidate in microsomal preparations from the developing cotyledons of safflower (*Carthamus tinctorius* L.) seed. *Biochem. J.*, **230**, 379-388.

Gurr M. I., 1974. — Biosynthesis of unsaturated fatty acids In *Biochemistry of lipids*., **4**, Kornberg H. L. and Phillips, D. C., ed., International Review of Science. Butterworth, London, 181-235.

Hammond E. G. and Fehr W. R., 1983. — Registration of A5 germplasm line of soybean (Reg no GP44). *Crop Sci.*, **23**, 192.

Hammond E. G. and Fehr W. R., 1984. — Improving the fatty acid composition of soybean oil. *J. Am. Oil Chem. Soc.*, **61**, 1713-1715.

Harwood J. L., 1984. — Effects of the environment on the acyl lipids of algae and higher plants. In *Structure, function and metabolism of plant lipids*. Siegenthaler P. A. and Eichenberger W., ed., Elsevier Science Publishers, Amsterdam, 543-550.

Hawke J. C. and Stumpf P. K., 1980. — Desaturation of oleic and linoleic acids by leaves of dark and light-growth maize seedlings. *Plant Physiol.*, **65**, 1027-1030.

Hawkins S. E., Fehr W. R. and Hammond, E. G., 1983*a*. — Resource allocation on breeding for fatty acid composition of soybean oil. *Crop Sci.*, **23**, 900-904.

Hawkins S. E., Fehr W. R., Hammond E. G. and Rodriquez de Cianzio S.. 1983*b*. — Use of tropical environments in breeding for oil composition of soybean genotypes adopted to temperate climates. *Crop Sci.*, **23**, 897-900.

Hazel J. R. and Prosser C. L., 1974. — Molecular mechanisms of temperature compensation on poikilotherms. *Physiol. Rev.*, **54**, 620-677.

Heemserk J. W. M. and Wintermans J. F. G. M., 1987.

— Role of the chloroplast in leaf acyl-lipid synthesis. *Physiol. Plantarum*, **70**, 558-568.

Heinz E., 1977. — Enzymatic reactions in galactolipid biosynthesis. In *Lipids and lipid polymers in higher plants*. Tevini M. and Lichtenthäler H. K., ed., Springer-Verlag, Berlin, 102-120.

Heinz E., Siebertz H. P., Linscheid M., Joyard J. and Douce R., 1979. — Investigation on the origin of diglyceride diversity in leaf lipids. In *Advances in the biochemistry and physiology of plant lipids*. Appelquist, L.-A. and Liljenberg C., ed., Elsevier, Amsterdam, 99-120.

Heinz E. and Roughan P. G., 1983. — Similarities and differences in lipid metabolism of chloroplasts isolated from 18:3 and 16:3 plants. *Plant Physiol.*, **72**, 273-279.

Heinz E. and Roughan P. G., 1982. — *De novo* synthesis, desaturation and acquisition of monogalactosyl diacylglycerol by chloroplasts from "16:3" and "18:3" plants. In *Biochemistry and metabolism of plant lipids*, **8**, Wintermans J. F. G. M. and Kuiper P. J. C., ed., Elsevier Biomedical Press, Amsterdam, 169-182.

Hilditch T. P. and Williams P. N., 1964. — *The chemical constitution of natural fats*. Chapman and Hall, London.

Hilton J. L., St. John J. B., Christiansen M. N. and Norris K. H., 1971. — Interaction of lipoidal materials and a pyridazinone inhibitor of chloroplast development. *Plant* Physiol., **48**, 171-175.

Howell R. W., Brim C. A. and Rinne R. W., 1972. — The plant geneticist contribution toward changing lipid and amino acid composition of soybean. *J. Am. Oil. Chem. Soc.*, **49**, 30-32.

Husemann W., Randwan S. S., Mangold H. K. M. and Bart W., 1980. — The lipids in photoautotrophic and heterotrophic cell suspension cultures of *Chenopodium rubyum*. *Planta*, **147**, 379-383.

James A. T. and Nichols B. W., 1966. — Lipids of photosynthetic system. *Nature*, **210**, 372-375.

Jamieson G. R. and Reid E. H., 1976. — The occurrence of hexadeca-7, 10, 13-trienoic acid in the leaf lipids of angiosperms. *Phytochemistry*, **10**, 1837-1843.

Jaworski J. G., 1987. — Biosynthesis of monoenoic and polyenoic fatty acids. In *The biochemistry of plants*, **9**, Stumpf P. K. and Conn E. E., ed., Academic Press, Orlando, 159-174.

Jaworski J. G. and Stumpf P. K., 1974. — Properties of a soluble stearyl-ACP desaturase from maturing *Carthamus tinctorius*. *Arch. Biochem. Biophys.*, **162**, 158-165.

Joyard J. and Douce R., 1977. — Site of synthesis of phosphatidic acid and diacylglycerol in spinach chloroplasts. *Biochim. Biophys. Acta.*, **486**, 273-285.

Khan M. U., Chapman D. J., Lem N. W., Chandorkar K. R. and Williams J. P., 1979 *a*. Effects of substituted pyridazinones. In *Advances in the biochemistry and physiology of plant lipids*. Appelquist, L.-A. and Liljenberg C., ed., Elsevier, Amsterdam, 415-420.

Khan M. U., Lem N. W., Chandorkar K. R. and Williams J. P., 1979 *b*. — Effects of substituted pyridazinones (San 6706, San 9774, San 9785) on glycerolipids and their associate fatty acids in leaves of *Vicia faba* and *Hordeum vulgare*. *Plant Physiol.*, **64**, 300-305.

Knauf V. C., 1987. — The application of genetic engineering to oil crops. *Trends Biotech.*, **5**, 40-47.

Kuiper P. J. C., 1984. — Lipid metabolism of higher plants as a factor in environmental adaptation. In *Structure, function and metabolism of plant lipids*. Siegenthaler P. A. and Eichenberger, W., ed., Elsevier Science Publishers, Amsterdam, 51-54.

Laskay G., Fardas T., Lehoczki E., and Gulya K., 1983. — Effects of pyridazinone herbicides during chloroplast development in detached barley leaves. II. Effects on lipid content, fatty acid composition and ultrastructure of chloroplasts. *Z. Naturforsch.*, **38**, 741-747.

Leech R. M. and Murphy D. J., 1976. - The cooperative function of chloroplasts in the synthesis of small molecules in the intact chloroplast. In *The intact chloroplast*. Barber J., ed. Elsevier-Holland, Tonn, 365.

Leech R. M. and Walton C. A., 1983. — Modification of fatty acid composition during chloroplast ontogeny and the effects on thylakoids appression and primary photochemistry. In *Biosynthesis and function of plant lipids*. Thompson W. W., Mudd J. B. and Gibbs M., ed., Waverly Press, Baltimore, 56-80.

Lem N. W. and Williams J. P., 1981. — Desaturation of fatty acids associated with monogalactosyl diacylglycerol: The effects of San 6706 and San 9785. *Plant Physiol.*, **68**, 944-949.

MacCarthy J. M. and Stumpf P. K., 1980. — The effects of different temperature on fatty acid synthesis and polyunsaturation in cell suspension culture. *Planta*, **147**, 389-392.

Martin B. A. and Rinne R. W., 1985. — Relationship between fatty acid composition of vegetative and reproductive structure of six soybean genotypes. *Crop Sci.*, **25**, 1055-1058.

McKeon T. A. and Stumpf P. K., 1982. — Purification and characterization of the stearoyl-acyl carrier protein desaturase and the acyl-acyl carrier protein thioesterase from maturing seeds of safflower. *J. Biol. Chem.*, **25**, 12141-12147.

Moore T. S. J., 1982. — Phospholipid biosynthesis. *Annu. Rev. Plant Physiol.*, **33**, 235-259.

Murata N. and Yamaya J., 1984. — Temperature dependent phase behavior of phosphatidylglycerols from chilling-sensitive and chilling-resistant plants. *Plant Physiol.*, **74**, 1016-1024.

Murphy D. J., Harwood J. L., Lee K. A., Roberto F., Stumpf P. K. and St. John J. B. 1985. — Differential responses of a range of photosynthetic tissues to a substituted pyridazinone, Sandoz 9785, specific effects on fatty acid desaturation. *Phytochemistry*, **24**, 1923-1929.

Murphy D. J., Harwood J. L., St. John J. B. and Stumpf, P. K., 1980. — Effect of a substituted pyridazinone compound BASF 13-338 on membrane lipid synthesis in photosynthetic tissues. *Biochem. Soc. Trans.*, **8,** 119-120.

Murphy D. J. and Stumpf P. K., 1979. — Light dependent induction of polyunsaturated fatty acid biosynthesis in greening cucumber cotyledons. *Plant Physiol.*, **63,** 328-335.

Murphy D. J., Woodrow J. E. and Mukherjee K. D., 1985. — Substrate specificities of the enzymes of the oleate desaturase system from photosynthetic tissue. *Biochem. J.*, **225,** 267-270.

Nagai J. and Bloch K., 1968. — Enzymatic desaturation of stearyl-ACP. *J. Biol. Chem.*, **243,** 4626-4633.

Nichols B. W., 1965. — Light induced changes in lipids composition during greening of etiolated pea seedlings. *Plant Physiol.*, **47,** 329-334.

Norman H. A. and St. John J. B., 1986. — Metabolism of unsaturated monogalactosyldiacylglycerol molecular species in *Arabidopsis thaliana* reveals different sites and substrates for linolenic acid synthesis. *Plant Physiol.*, **81,** 731-736.

Norman H. A. and St. John J. B., 1987. — Differential effects of a substituted pyridazinone, BASF 13-338, on pathways of monogalactosyl diacylglycerol synthesis in *Arabidopsis. Plant Physiol.*, **85,** 684-688.

Ohlrogge J. B., Shine W. E. and Stumpf P. K., 1978. — Characterization of plant acyl-ACP and acyl-CoA hydrolases. *Arch. Biochem. Biophys.*, **189,** 382-391.

Ohnishi J. and Yamada M., 1982. — Glycerolipid synthesis in *Avena* leaves during greening of etiolated seedlings: III. Synthesis of alpha-linoleoyl-monogalactosyl diacylglycerol from liposomal linoleoyl-phosphatidylcholine by *Avena* plastids in the presence of phosphatidylcholine exchange protein. *Plant Cell Physiol.*, **23,** 767-773.

Oquist G. and Liljenber, C., 1981. — Lipid and fatty acid composition of chloroplast thylakoids isolated from *Betula pendula* leaves in different stages of development of acclimated quantum flux densities. *Z. Pflanzenphysiol.*, **104,** 233-243.

Ory R. L., St. Angelo A. V., Conkerton E. J. and Chapital D. C., 1984. — Properties of peanuts (*Arachis hypogaea* L.) from bioregulator-treated plants. In *Bioregulators: chemistry and uses, ACS.* Ory R. L. and Rittig F. R., ed., Symposium Series, **25,** 83-88.

Ouedraogo M., Trémolières A. and Hubac C., 1984. — Change in fatty acids composition during water stress in cotton plants. Relation with drought resistance induced by far red light. *Z. Pflanzenphysiol.*, **114,** 239-245.

Oursel A., Escoffier A., Kader J. C., Dubacq J. P. and Trémolières A., 1987. — Last step in the cooperative pathway for galactolipid synthesis in spinach leaves: formation of monogalactosyldiacylglycerol with C18 polyunsaturated acyl groups at both carbon atoms of the glycerol. *FEBS Lett.*, **219,** 393-399.

Pham-Thi A. T., Borrel-Flood C., Vierra da Silva J., Justin A. M. and Mazliak P., 1987. — Effects of drought on [1-^{14}C]-oleic and [1-^{14}C]-linoleic acid desaturation in cotton leaves. *Physiol. Plantarum*, **69,** 147-150.

Pham-Thi A. T., Borrel-Flood C., Vierra da Silva J., Justin A. M. and Mazliak P., 1985. — Effects of water stress on lipid metabolism in cotton leaves. *Phytochemistry*, **24,** 723-727.

Pham-Thi A. T., Borrel-Flood C. and Vierra da Silva J., 1982. — Effects of water stress on lipid and fatty acid composition of cotton leaves. In *Biochemistry and metabolism of plant lipids.* Wintermans J. F. G. M. and Kuiper P. J. C., ed., Elsevier, Amsterdam, 451-454.

Quinn P. J. and Williams W. P., 1983. — The structural role of lipids in photosynthetic membranes. *Biochim. Biophys. Acta*, **737,** 223-266.

Raison J., 1980. — Membrane lipids: structure and function. In *The Biochemistry of Plants*, **4,** Stumpf P. K. and Conn E. E., ed., Academic Press, New York. 57-83.

Rakow G., 1973. — Selektion auf Linol- und linolensauregehalt in Rapssaman nach mutagener Benhandlung, *Z. Pflanzenzuchtg.*, **69,** 62.

Rébeillé F., Bligny R. and Douce R., 1980. — Roles of oxygen and temperature in the fatty acid composition of isolated sycamore (*Acer pseudoplatanus* L.) cells. *Biochim. Biophys. Acta.*, **620,** 1-9.

Reubel A., Rinne R. W. and Canvin D. T., 1972. — Protein, oil, and fatty acid in developing soybean seeds. *Crop Sci.*, **12,** 739-741.

Robbelen G. A. and Nitch A., 1975. — Genetische and physiologische Untersuchungen neuer Mutanten. *Z. Pflanzenzuchtg.*, **75,** 93-98.

Roehm J. N. and Privett O. S., 1970. — Changes in the structure of soybean triglycerides during maturation. *Lipids.*, **5,** 353-358.

Rochester C. P. and Bishop D. G., 1984. — The role of lysophosphatidylcholine in lipid synthesis by developing sunflower (*Helianthus annus* L) seed microsomes. *Arch. Biochem. Biophys.*, **232,** 249-258.

Roughan P. G., 1975. — Phosphatidylcholine: donor of 18-carbon unsaturated fatty acids from glycerolipid biosynthesis. *Lipids*, **10,** 609-614.

Roughan P. G., 1985. — Phosphatidylglycerol and chilling sentitivity in plants. *Plant Physiol.*, **77,** 740-746.

Roughan P. G., Holland R. and Slack C. R., 1979 *a.* — On the control of long chain fatty acid synthesis in isolated intact spinach (*Spinacia oleracea*) chloroplasts. *Biochem. J.*, **184,** 193-202.

Roughan P. G., Holland R. and Slack C. R., 1980. — The role of chloroplasts and microsomal fractions in polar lipid synthesis from [1-^{14}C]-acetate by cell free preparation from spinach (*Spinacia oleracea*) leaves. *Biochem. J.*, **188,** 17-24.

Roughan P. G., Mudd J. B., McManus T. T. and Slack C. R., 1979*b*. — Linoleate and linolenate synthesis by isolated spinach (*Spinacia oleracea*) chloroplasts. *Biochem. J.*, **184**, 378-381.

Roughan P. G. and Slack C. R., 1980. — The role of chloroplast in leaf lipid metabolism and polyunsaturated fatty acid synthesis. In *Biogenesis and function of plant lipids*. Mazliak P., Benveniste P., Coste C. and Douce R., ed., Elsevier/North Holand, Amsterdam, 11-18.

Roughan P. G. and Slack C. R., 1982. — Cellular organization of glycerolipid metabolism. *Annu. Rev. Plant Physiol.*, **33**, 97-132.

Roughan P. G. and Slack C. R., 1984. — Glycerolipid synthesis in leaves. *Trends Biochem. Sci.*, **9**, 383-386.

Sauer A. and Heise K. P., 1982. — Correlation between glycerol-3-phosphate uptake and lipid synthesis in spinach chloroplasts. In *Biochemistry and metabolism of plant lipids*, **8**, Wintermans J. F. and Kuiper P. J. C., ed., Elsevier Biomedical Press, Amsterdam, 187-190.

Sauer A. and Robinson D. G., 1985. — Subcellular localization of enzymes involved in lecithin biosynthesis in maize roots. *J. Exp. Bot.*, **36**, 1257-1266.

Siebertz H. P., Heinz S., Joyard J. and Douce R., 1980. — Labelling *in vitro* and *in vivo* of molecular species of lipids from chloroplast envelopes and thylakoids. *Eur. J. Biochem.*, **108**, 177-185.

Slack C. R. and Roughan P. G., 1978. — Rapid temperature-induced changes in the fatty acid composition of certain lipids in developing linseed and soybean cotyledons. *Biochem. J.*, **170**, 437-439.

Slack C. R., Roughan P. G. and Balasingham N., 1978. — Labelling of glycerolipids in the cotyledons of developing oilseeds by [1-^{14}C]acetate and [2-^{3}H]glycerol. *Biochem. J.*, **170**, 421-433.

Slack C. R., Roughan P. G. and Browse J. A., 1979. — Evidence for an oleoyl-phosphatidylcholine desaturase in microsomal preparations from cotyledons of safflower seeds. *Biochem. J.*, **179**, 649-656.

Slack C. R., Roughan P. G. and Terpstra J., 1976. — Some properties of a microsomal oleate desaturase from leaves. *Biochem. J.*, **155**, 71-80.

Smith K. J., 1981. — Improving the quality of the soybean. *J. Am. Oil Chem. Soc.*, **58**, 135-139.

Smouse T. H., 1979. — A review of soybean oil reversion flavor. *J. Am. Oil Chem. Soc.*, **56**, 747-751.

Somerville C. R., McCourt P., Kunst L. and Browse J., 1987. — Mutants of *Arabidopsis* deficient in fatty acid desaturation. In *The metabolism, structure, and function of plant lipids*. Stumpf P. K., Mudd J. B. and Nes W. D., ed., Plenum Press, New York, 683-688.

Sparace S. and Mudd J. B., 1982. — Phosphatidylglycerol synthesis in spinach chloroplasts: characterization of the new synthesized molecule. *Plant Physiol.*, **70**, 1260-1264.

St. John J. B., 1976. — Manipulation of galactolipid fatty acid composition with substituted pyridazinones. *Plant Physiol.*, **57**, 38-40.

St. John J. B. and Hilton J. L., 1976. — Structure versus activity of substituted pyridazinones as related to mechanism of action. *Weed Sci.*, **24**, 579-583.

St. John J. B., Christiansen M. N. and Terlizi D. E., 1984. — Chemical manipulation of soybean (*Glycine max* L. Merr.) seed quality. In *Bioregulators: chemistry and uses*. ACS Ory R. L. and Ritting I. R., ed., Symposium Series, **25**, 257-263.

Stobart A. K. and Stymne S., 1985. — The regulation of the fatty acid composition of the triacylglycerols in microsomal preparations from avocado mesocarp and the developing cotyledons of safflower. *Planta*, **163**, 119-125.

Stobart A., Stymne S. and Glad G., 1983. — The synthesis of linoleate and phosphatidic acid and its relationship to oil production in the microsome of developing seed of safflower (*Carthamus tinctorius* L. var. Gila). *Biochim. Biophys. Acta.*, **754**, 292-297.

Stumpf P. K., 1980. — Biosynthesis of saturated and unsaturated fatty acids. In *Biochemistry of plants, lipids: structure and function.*, **4**, Stumpf P. K. and Conn E. E., ed., Academic Press, New York, 177-204.

Stumpf P. K. and Porra R. J., 1976. — Lipid biosynthesis in developing and germinating soybean cotyledons. *Arch. Biochem. Biophys.*, **176**, 63-70.

Stymne S. and Appelquist L. A., 1978. — The biosynthesis of linoleate from oleoyl-CoA and phosphatidylcholine in microsomes of developing safflower seeds. *Eur. J. Biochem.*, **90**, 223-229.

Stymne S. and Appelquist L. A., 1980. — The biosynthesis of linoleate and alpha-linolenate in homogenates from developing soya bean cotyledons. *Plant Sci. Lett.*, **17**, 287-294.

Stymne S. and Stobart A. K., 1984*a*. — Evidence for the reversibility of the acyl-CoA: lysophosphatidyl-choline acyltransferase in microsomal preparations from developing safflower (*Carthamus tinctorius* L.) cotyledons and rat liver. *Biochem. J.*, **223**, 305-314.

Stymne S. and Stobart A. K. 1984*b*. — The biosynthesis of triacylglycerols in microsomal preparations of developing cotyledons of sunflower (*Helianthus annuus* L.) *Biochem. J.*, **220**, 481-488.

Stymne S. and Stobart A. K., 1985. — Oil synthesis *in vitro* in microsomal membranes from developing cotyledons of *Linum usitatissimum* L. *Planta*, **164**, 101-104.

Stymne S., Stobart A. K. and Glad G., 1983. — The role of the acyl-CoA pool in the synthesis of polyunsaturated 18-carbon fatty acids and triacylglycerol production in the microsomes of developing safflower seeds. *Biochim. Biophys. Acta*, **752**, 198-208.

Tchang F., Connan A., Robert D. and Trémolières A., 1985. — Effects of light and temperature on α-linolenic

acid biosynthesis in developing cotyledons of *Pharbitis nil*. *Physiol. Vég.*, **23**, 361-371.

Thies W., 1970. — Chloroplast development and biogenesis of linolenic acid in ripening cotyledons of rapeseed. *Proc. Intour. Rapeseed Congress Ste. Adele.*, 348.

Thompson G. A., 1980. — *The regulation of membrane lipid metabolism.* CRC Press, Boca Raton.

Trémolières A., 1985. — Light effects on membrane lipid metabolism in higher plants. *Physiol. Vég.*, **23**, 955-961.

Trémolières A., Drapier D., Dubacq J. P. and Mazliak P., 1980*a*. — Oleyl-coenzyme A metabolism by subcellular fractions from growing pea leaves. *Plant. Sci. Lett.*, **18**, 257-269.

Trémolières A., Dubacq J. P. and Drapier D., 1982. — Unsaturated fatty acids in maturing seeds of sunflower and rape: Regulation by temperature and light intensity. *Phytochemistry*, **21**, 41-45.

Trémolières A., Dubacq J.-P., Drapier D., Muller M. and Mazliak P., 1980*b*. — *In vitro* cooperation between plastids and microsomes in the biosynthesis of leaf lipids. *FEBS Lett.*, **114**, 135-138.

Trémolières A. and Lepage M., 1971. — Changes in lipid composition during greening of etiolated pea seedlings. *Plant Physiol.*, **47**, 329-279.

Trémolières A. and Mazliak P., 1974. — Biosynthesis of α-linolenic acid in developing pea leaves, *in vivo* and *in vitro* study. *Plant Sci. Lett.*, **2**, 193-201.

Trémolières A. T. and Mazliak P., 1970. — Formation des lipides au cours du développement de la feuille de Tréfle (*Trifolium repens* L.). *Physiol. Vég.*, **8**, 135-139.

Trémolières H., Trémolières A. and Mazliak P., 1978. — Effects of light and temperature on fatty acid desaturation during the maturation of rapeseed. *Phytochemistry*, **17**, 685-687.

Wang X. M. and Hildebrand D. F., 1987. — Effect of a substituted pyridazinone on the decrease of lipoxygenase activity in soybean cotyledons. *Plant Sci.*, **51**, 29-36.

Wang X. M., Hildebrand D. F. and Collins G. B., 1987*b*. — Identification of proteins associated with changes in the linolenate content of soybean cotyledons. In *The metabolism, structure, and fonction of plant lipids.* Stumpf P. K., Mudd J. B. and Nes W. D., ed., Plenum Press, New York, 533-535.

Wang X. M., Hildebrand D. F., Norman H. A. Dahmer M. L., St. John J. B. and Collins G. B., 1987*a*. — Reduction of linolenate content in soybean cotyledons by a substituted pyridazinone. *Phytochemistry*, **26**, 955-960.

Wang X.-M., Norman H. A., St. John J. B., Yin T. and Hildebrand D. F., 1988. — Decrease of linolenate levels in seeds and roots of low linolenate mutants in soybean. *Phytochemistry*, in press.

Wilcox J. R. and Cavins J. F., 1985. — Inheritance of low linolenic acid content of the seed oil of a mutant in *Glycine max*. *Theor. Appl. Genet.*, **71**, 74-78.

Wilcox J. R. and Cavins J. F., 1987. Gene symbol assigned for linolenic acid mutant in the soybean. *J. Hered.*, **78**, 410.

Wilcox J. R., Cavins J. F. and Nielsen N. C., 1984. — Genetic alteration of soybean oil composition by a chemical mutagen. *J. Am. Oil Chem. Soc.*, **61**, 97-100.

Willemot C., 1977. — Simultaneous inhibition of linolenic acid synthesis in winter wheat root and frost hardening by BASF 13-338, a derivative of pyridazinone. *Plant Physiol.*, **60**, 1-4.

Willemot C., Slack C. R., Browse J. and Roughan P. J., 1982. — Effect of BASF 13-338, a substituted pyridazinone, on lipid metabolism in leaf tissue of spinach, pea, linseed and wheat. *Plant Physiol.*, **70**, 78-81.

Williams J. P. and Khan M.-V., 1982. — Lipid biosynthesis in *Brassica napus*. I. ^{14}C-labelling kinetics of the fatty acids of the major glycerolipids. *Biochim. Biophys. Acta*, **713**, 177-184.

Williams J. P., Khan M.-V. and Mitchell K., 1983. — Galactolipid biosynthesis in leaves of 16:3 and 18:3 plants. In *Biosynthesis and function of plant lipids.* Thomson W. W., Mudd J. B. and Gibbs M., ed., Waverly Press, Baltimore, 28-39.

Wilson R. F., Burton J. W. and Brim C. A., 1981. — Progress in the selection for altered fatty acid composition by plant breeding. *Crop Sci.*, **21**, 788-791.

Wilson R. F., Weisinger H. H., Buck J. A. and Faulkner G. D., 1980. — Involvement of phospholipids in polyunsaturated fatty acid synthesis in developing soybean cotyledons. *Plant Physiol.*, **66**, 545-549.

Zarrouk M. and Cherif A., 1984. — Effect of sodium chloride on chloroplast lipid composition of olive leaves (*Olea europea* L.) In *Structure, function and metabolism of plant lipids.* Siegenthaler P. A. and Eichenberger W., ed., Elsevier Science Publishers, Amsterdam, 595-599.

Murphy D. J., Harwood J. L., St. John J. B. and Stumpf, P. K., 1980. — Effect of a substituted pyridazinone compound BASF 13-338 on membrane lipid synthesis in photosynthetic tissues. *Biochem. Soc. Trans.*, **8,** 119-120.

Murphy D. J. and Stumpf P. K., 1979. — Light dependent induction of polyunsaturated fatty acid biosynthesis in greening cucumber cotyledons. *Plant Physiol.*, **63,** 328-335.

Murphy D. J., Woodrow J. E. and Mukherjee K. D., 1985. — Substrate specificities of the enzymes of the oleate desaturase system from photosynthetic tissue. *Biochem. J.*, **225,** 267-270.

Nagai J. and Bloch K., 1968. — Enzymatic desaturation of stearyl-ACP. *J. Biol. Chem.*, **243,** 4626-4633.

Nichols B. W., 1965. — Light induced changes in lipids composition during greening of etiolated pea seedlings. *Plant Physiol.*, **47,** 329-334.

Norman H. A. and St. John J. B., 1986. — Metabolism of unsaturated monogalactosyldiacylglycerol molecular species in *Arabidopsis thaliana* reveals different sites and substrates for linolenic acid synthesis. *Plant Physiol.*, **81,** 731-736.

Norman H. A. and St. John J. B., 1987. — Differential effects of a substituted pyridazinone, BASF 13-338, on pathways of monogalactosyl diacylglycerol synthesis in *Arabidopsis. Plant Physiol.*, **85,** 684-688.

Ohlrogge J. B., Shine W. E. and Stumpf P. K., 1978. — Characterization of plant acyl-ACP and acyl-CoA hydrolases. *Arch. Biochem. Biophys.*, **189,** 382-391.

Ohnishi J. and Yamada M., 1982. — Glycerolipid synthesis in *Avena* leaves during greening of etiolated seedlings: III. Synthesis of alpha-linoleoyl-monogalactosyl diacylglycerol from liposomal linoleoyl-phosphatidylcholine by *Avena* plastids in the presence of phosphatidylcholine exchange protein. *Plant Cell Physiol.*, **23,** 767-773.

Oquist G. and Liljenber, C., 1981. — Lipid and fatty acid composition of chloroplast thylakoids isolated from *Betula pendula* leaves in different stages of development of acclimated quantum flux densities. *Z. Pflanzenphysiol.*, **104,** 233-243.

Ory R. L., St. Angelo A. V., Conkerton E. J. and Chapital D. C., 1984. — Properties of peanuts (*Arachis hypogaea* L.) from bioregulator-treated plants. In *Bioregulators: chemistry and uses, ACS.* Ory R. L. and Rittig F. R., ed., Symposium Series, **25,** 83-88.

Ouedraogo M., Trémolières A. and Hubac C., 1984. — Change in fatty acids composition during water stress in cotton plants. Relation with drought resistance induced by far red light. *Z. Pflanzenphysiol.*, **114,** 239-245.

Oursel A., Escoffier A., Kader J. C., Dubacq J. P. and Trémolières A., 1987. — Last step in the cooperative pathway for galactolipid synthesis in spinach leaves: formation of monogalactosyldiacylglycerol with C18 polyunsaturated acyl groups at both carbon atoms of the glycerol. *FEBS Lett.*, **219,** 393-399.

Pham-Thi A. T., Borrel-Flood C., Vierra da Silva J., Justin A. M. and Mazliak P., 1987. — Effects of drought on [1-^{14}C]-oleic and [1-^{14}C]-linoleic acid desaturation in cotton leaves. *Physiol. Plantarum*, **69,** 147-150.

Pham-Thi A. T., Borrel-Flood C., Vierra da Silva J., Justin A. M. and Mazliak P., 1985. — Effects of water stress on lipid metabolism in cotton leaves. *Phytochemistry*, **24,** 723-727.

Pham-Thi A. T., Borrel-Flood C. and Vierra da Silva J., 1982. — Effects of water stress on lipid and fatty acid composition of cotton leaves. In *Biochemistry and metabolism of plant lipids.* Wintermans J. F. G. M. and Kuiper P. J. C., ed., Elsevier, Amsterdam, 451-454.

Quinn P. J. and Williams W. P., 1983. — The structural role of lipids in photosynthetic membranes. *Biochim. Biophys. Acta*, **737,** 223-266.

Raison J., 1980. — Membrane lipids: structure and function. In *The Biochemistry of Plants*, **4,** Stumpf P. K and Conn E. E., ed., Academic Press, New York. 57-83.

Rakow G., 1973. — Selektion auf Linol- und linolensaure-gehalt in Rapssaman nach mutagener Benhandlung. *Z. Pflanzenzuchtg.*, **69,** 62.

Rébeillé F., Bligny R. and Douce R., 1980. — Roles of oxygen and temperature in the fatty acid composition of isolated sycamore (*Acer pseudoplatanus* L.) cells. *Biochim. Biophys. Acta.*, **620,** 1-9.

Reubel A., Rinne R. W. and Canvin D. T., 1972. — Protein, oil, and fatty acid in developing soybean seeds. *Crop Sci.*, **12,** 739-741.

Robbelen G. A. and Nitch A., 1975. — Genetische and physiologische Untersuchungen neuer Mutanten. *Z. Pflanzenzuchtg.*, **75,** 93-98.

Roehm J. N. and Privett O. S., 1970. — Changes in the structure of soybean triglycerides during maturation. *Lipids.*, **5,** 353-358.

Rochester C. P. and Bishop D. G., 1984. — The role of lysophosphatidylcholine in lipid synthesis by developing sunflower (*Helianthus annus* L) seed microsomes. *Arch. Biochem. Biophys.*, **232,** 249-258.

Roughan P. G., 1975. — Phosphatidylcholine: donor of 18-carbon unsaturated fatty acids from glycerolipid biosynthesis. *Lipids*, **10,** 609-614.

Roughan P. G., 1985. — Phosphatidylglycerol and chilling sentitivity in plants. *Plant Physiol.*, **77,** 740-746.

Roughan P. G., Holland R. and Slack C. R., 1979*a.* — On the control of long chain fatty acid synthesis in isolated intact spinach (*Spinacia oleracea*) chloroplasts. *Biochem. J.*, **184,** 193-202.

Roughan P. G., Holland R. and Slack C. R., 1980. — The role of chloroplasts and microsomal fractions in polar lipid synthesis from [1-^{14}C]-acetate by cell free preparation from spinach (*Spinacia oleracea*) leaves. *Biochem. J.*, **188,** 17-24.

Roughan P. G.,
C. R., 1979*b.*
isolated spina
chem. J., **184,**
Roughan P. G.
chloroplast in
ted fatty acid
plant lipids. M
Douce R., ed
11-18.
Roughan P. G.
organization
Plant Physiol.
Roughan P. G.
synthesis in le
Sauer A. and H.
glycerol-3-phc
spinach chlor
of plant lipids,
ed., Elsevier B
Sauer A. and
localization of
in maize root
Siebertz H. P.,
— Labelling
of lipids from
Eur. J. Bioch
Slack C. R. and
ture-induced
certain lipids
ledons. *Bioch*
Slack C. R., Ro
— Labelling
loping oilseed
Biochem. J., 1
Slack C. R., Ro
Evidence for
in microsom
wer seeds. *Bi*
Slack C. R., Rou
properties of
leaves. *Bioch*
Smith K. J., 198
J. Am. Oil C
Smouse T. H., 1
flavor. *J. Am*
Somerville C.
1987. — Mu
desaturation
of plant lipid
ed., Plenum
Sparace S. and
synthesis in
the new synt
1264.

Acknowledgment

This work was supported in part by Grant GM-18278 from the National Institutes of Health. Dr. A. A. Khan contributed to some of the methods described.

[34] Stearoyl-Acyl Carrier Protein Desaturase from Safflower Seeds

By TOM MCKEON and PAUL K. STUMPF

Stearoyl-ACP (acyl carrier protein) desaturase is the enzyme responsible for the synthesis of oleic acid in plants. Nagai and Bloch, who first characterized the activity, found that the enzyme requires stearoyl-ACP, reduced ferredoxin, and molecular oxygen.[1,2]

$$\text{Stearoyl-ACP} + \text{ferredoxin(II)} + O_2 + 2H^+ \rightarrow \text{oleoyl-ACP} + \text{ferredoxin(III)} + 2H_2O$$

The stearoyl-ACP desaturase is easily extracted into buffer without the use of detergents, has no requirement for added lipid, and has a lipid-insoluble substrate,[1] all in marked contrast to the stearoyl-CoA desaturase of animal systems.[3] However, because the plant and animal desaturases both require oxygen and an electron transfer system to carry out the same chemical reaction, it is thought that the mechanism of the reaction may be the same for both types of enzyme.

Nagai and Bloch found the stearoyl-ACP desaturase in photosynthetic tissue—*Euglena gracilis* and spinach chloroplasts.[1,2] Subsequently, Jaworski and Stumpf characterized the activity in immature safflower (*Carthamus tinctorius*) seed,[4] a nonphotosynthetic tissue. The activity is also present in avocado mesocarp,[5] immature soybean cotyledons,[6] immature jojoba nuts,[7] and immature coconut.[8] However, this report describes only the stearoyl-ACP desaturase from safflower.

[1] J. Nagai and K. Bloch, *J. Biol. Chem.* **241,** 1925 (1966).
[2] J. Nagai and K. Bloch, *J. Biol. Chem.* **243,** 4626 (1968).
[3] P. W. Holloway, this series, Vol. 35 [31].
[4] J. G. Jaworski and P. K. Stumpf, *Arch. Biochem. Biophys.* **162,** 158 (1974).
[5] B. S. Jacobson, J. G. Jaworski, and P. K. Stumpf, *Plant Physiol.* **54,** 484 (1974).
[6] P. K. Stumpf and R. J. Porra, *Arch. Biochem. Biophys.* **176,** 63 (1976).
[7] M. R. Pollard, T. McKeon, L. M. Gupta, and P. K. Stumpf, *Lipids* 651 (1979).
[8] T. McKeon, unpublished data.

ISBN 0-12-181971-X

Assay Method

Principle. The assay for stearoyl-ACP desaturase is based on the measurement of ^{14}C-labeled oleic acid produced by desaturation of ^{14}C-labeled stearoyl-ACP. Separation and quantitation of the ^{14}C-labeled fatty acids are carried out by thin-layer chromatography and scintillation counting or by gas–liquid chromatography and radioactive counting in a proportional counter.

Reagents

PIPES, 0.10 *M*, pH 6.0
NADPH, 25 m*M*, freshly prepared in 0.1 *M* Tricine, pH 8.2
Bovine serum albumin (BSA), lipid free, 10 mg/ml in water
Dithiothreitol (DTT), 0.10 *M*, freshly prepared
Ferredoxin, 2 mg/ml (Sigma, spinach, type III)
NADPH: ferredoxin oxidoreductase (Sigma), 2.5 units/ml
Catalase (Sigma, bovine liver, 800,000 units/ml)
[^{14}C]Stearoyl-ACP, 10 μM in 0.1 *M* PIPES, pH 5.8; its synthesis is described after the assay procedure
NaOH, 8 *M*
H_2SO_4, 4 *M*
Stearic acid and oleic acid, 1 mg/ml each in acetone
Petroleum ether
Diazomethane (20 mg/ml in diethyl ether)
$AgNO_3$–silica gel G thin-layer plates, 0.25 mm thick (Redi-Coat AG, Supelco)
2,7-Dichlorofluorescein, 0.1% in methanol

Procedure. The following reagents are added for each assay: water, 150 μl; DTT 5 μl; BSA, 10 μl; NADPH, 15 μl; ferredoxin, 25 μl; NADPH : ferredoxin oxidoreductase, 3 μl; and catalase, 1 μl. This mixture is kept at room temperature for 10 min and is then added to a 13 × 100 mm screw-cap test tube containing 250 μl of PIPES buffer. The stearoyl-ACP desaturase preparation is added in a volume of 10 μl, and the reaction is started by adding 30 μl of stearoyl-ACP and incubating at 23° with shaking for 10 min. The reaction is stopped by adding 125 μl of 8 *M* NaOH and 0.1 ml of the fatty acid solution. The tubes are capped and incubated for 1 hr at 80°. The mixture is acidified with 160 μl of 4 *M* H_2SO_4 and vigorously extracted three times with 2-ml portions of petroleum ether. The extract is evaporated under nitrogen, methylated with 0.5 ml of diazomethane solution for 30 min on ice, and then evaporated to dryness. The methyl esters of stearate and oleate are then separated and quantitated by either of two methods: thin-layer chromatography on $AgNO_3$–silica gel plates as described by Holloway[3] or gas chromatography (10%

DEGS-PS on Supelcoport 80/10; 6 ft. × $\frac{1}{4}$ in. column at 180°) followed by counting of the radioactivity in a gas proportional counter. Radio–gas chromatography avoids the slight complication of correcting for the [^{14}C]palmitate contaminant present in most stearoyl-ACP preparations, but, for accuracy and sensitivity, thin-layer chromatography is the method of choice. One unit of activity is defined as 1 μmol of oleate produced per milligram of protein per minute.

An alternative method for reducing ferredoxin uses a chloroplast grana suspension, ascorbic acid, 2,6-dichlorophenolindophenol, and light. This system has been described in detail by Jaworski and Stumpf.[4]

Stearoyl-ACP Synthesis

Procedure. Stearoyl-ACP is made with a safflower fatty acid synthase system, [^{14}C]malonyl-CoA, and *Escherichia coli* ACP. The method described herein differs only slightly from that described by Jaworski and Stumpf.[9]

Immature safflower seeds are suspended in an equal volume of 0.10 *M* potassium phosphate, 5 m*M* sodium ascorbate, pH 6.8, and are homogenized with a Polytron instrument for three half-minute periods at half speed, centrifuged at 12,000 *g* for 20 min, and filtered through four layers of cheesecloth and one layer of Miracloth. The safflower supernatant is used as a source of fatty acid synthase with no further purification. It is stable when frozen for 6 weeks.[8]

The incubation medium contains the following components in a total volume of 5 ml: water 2.2 ml; 25 m*M* NADH, 100 μl; 25 m*M* NADPH, 100 μl; 1.0 *M* Tricine (K^+), pH 7.9, 250 μl; 0.10 *M* DTT 25 μl; 200 m*M* $MgCl_2$, 25 μl; 1.0 m*M* malonyl-CoA, 1.0 ml; [1,3-^{14}C]malonyl-CoA (50–60 mCi/mmol), 10 μCi in 500 μl; and ACP (4 mg/ml), 175 μl. The ACP used is purified from *E. coli* by the method of Majerus *et al.*[10] to 90% purity, and is reduced with 1 m*M* DTT for 15 min just prior to use. The reaction mixture is carefully bubbled with nitrogen for 5 min; 625 μl of safflower supernatant are added, then the mixture is again bubbled with nitrogen for a minute, stoppered, and placed in a 23° water bath. The reaction is stopped after 45 min by the addition of 0.55 ml of 50% trichloroacetic acid (TCA) in the hood and bubbled with nitrogen to displace $^{14}CO_2$; it is held on ice for 30 min and centrifuged at 5000 *g* for 5 min. The pellet is redissolved in 2.5 ml of 0.10 *M* PIPES, pH 5.8, titrating with 1 *M* KOH if necessary; debris is removed by centrifugation, and solid ammonium sulfate is added

[9] J. G. Jaworski and P. K. Stumpf, *Arch. Biochem. Biophys.* **162**, 166 (1974).
[10] P. W. Majerus, A. W. Alberts, and P. R. Vagelos, this series, Vol. 14 [6].

to 70% saturation (0°). The precipitate is centrifuged at 12,000 g for 10 min, and the supernatant is acidified with 50% TCA to 10%. The TCA precipitate is dissolved in 1 ml of PIPES buffer as before; insoluble material is removed by centrifugation, and the concentration of the stearoyl-ACP is adjusted to 10 μM. This preparation provides 25–40% of the theoretical yield of ^{14}C in acyl-ACP. The product, as analyzed by radio-gas chromatography and $AgNO_3$–silica gel TLC, contains 80–90% stearoyl-ACP, 10–20% palmitoyl-ACP, and less than 0.5% oleoyl-ACP. Frozen solutions of acyl-ACP are stable for over 2 months.

An alternative method for making acyl-ACP of specific chain length is the acyl-ACP synthetase reaction described by Spencer *et al.*[11] Since a specific fatty acid may be ligated to the ACP with this sytem, it has been used to make the substrates employed in specificity studies. However, this system does not efficiently ligate stearic acid to ACP (2–4%) in our hands; therefore, the fatty acid synthase reaction is routinely used to produce substrate for desaturase assays. Another method for making acyl-ACP is described in this volume [21].

Purification

Materials

Immature safflower seed, Gila variety, harvested at approximately 14–18 days after flowering, as indicated by a charcoal gray seed coat

Acetone, reagent grade, −20°

Diethyl ether, anhydrous

DEAE-cellulose, equilibrated with 0.02 M potassium phosphate, pH 6.8

ACP-Sepharose 4B, 2 mg of ACP per milliliter of wet gel; the column material was made with purified *E. coli* ACP, and cyanogen bromide activated Sepharose 4B by the method of March *et al.*[12] The reaction was carried out at pH 6.5 in 0.1 M $NaHCO_3$ for 1 day at 4°, and 70% of the ACP was covalently bound to the Sepharose

Potassium phosphate buffers, 0.02 M, 0.10 M, and 0.30 M, all pH 6.8, sterilized and degassed

Procedure

Acetone Powder. Immature safflower seeds (stored at −20°) are ground with an equal volume of acetone at high speed in a blender. The suspen-

[11] A. K. Spencer, A. D. Greenspan, and J. E. Cronan, Jr., *FEBS Lett.* **101**, 253 (1979).

[12] S. C. March, I. Parikh, and P. Cuatrecasas, *Anal. Biochem.* **60**, 149 (1974).

sion is suction-filtered, and the retained material is repeatedly extracted with acetone as above until the filtrate is clear and colorless. After the third extraction, the acetone suspension is passed through a coarse sieve to remove fragments of seed coat. Generally, five extractions are required to remove the lipids and phenolics. After the final filtration, the retained material is rinsed several times with a small volume of ether at $-20°$ to remove acetone, and then is kept under suction or in a vacuum desiccator to remove the last trace of ether. Stearoyl-ACP desaturase is stable in frozen seeds for at least 2 years and in the acetone powder for at least 3 months.

The following steps are carried out at 0–4°.

Acetone Powder Extract. Acetone powder from a given weight of seed is triturated with twice that weight of 0.02 *M* phosphate buffer and gently agitated for 1 hr. The suspension is then centrifuged at 12,000 *g* for 20 min and filtered through Miracloth, the supernatant, which contains the desaturase, is immediately applied to DEAE-cellulose or frozen. The activity in this preparation is stable for 3–4 weeks at $-20°$ or for 1 week at 4°.

DEAE-Cellulose Pass-through. Acetone powder extract is passed through a column of DEAE-cellulose (1 ml bed volume/3 ml extract), and the column is washed with one bed volume of 0.02 *M* phosphate buffer. The pass-through and effluent from the wash are collected. While this step does afford some purification (see the table), its principal purpose is to eliminate an acyl-ACP thioesterase present in the extract. Approximately 80% of the thioesterase is thus eliminated.[8]

ACP-Sepharose 4B column. The capacity of the ACP-Sepharose is 5 ml of DEAE-cellulose pass-through per milliliter of column material. Decreasing this ratio does not improve the percentage yield, and increasing the ratio decreases the percentage yield.

A column with a 20-ml bed volume is loaded at a flow rate of 0.5 ml per

PURIFICATION OF STEAROYL-ACYL CARRIER PROTEIN (ACP) DESATURASE

Step	Total protein[a] (mg)	Total activity (mU)	Specific activity (mU/mg)	Yield (%)	Purification factor
Acetone powder extract	380	205	0.55	—	—
DEAE-cellulose pass-through	170	162	0.95	79	1.7
ACP-Sepharose 4-B	0.34	38	110	19	200

[a] Protein was determined by the method of O. H. Lowry, N. J. Rosebrough, A. L. Farr, and R. J. Randall, *J. Biol. Chem.* **193**, 265 (1951), using bovine serum albumin as a standard.

minute, washed with two bed volumes of 0.02 *M* buffer and three bed volumes of 0.10 *M* buffer. The stearoyl-ACP desaturase is eluted with 0.30 *M* buffer and collected in fractions of 1.5 ml. The early fractions contain most of the contaminating acyl-ACP thioesterase; the most purified fractions of desaturase contain acyl-ACP thioesterase as 5–10% of the bulk protein.[8] The desaturase activity from this preparation is stable for 1 week at 4°.

Purity. The most purified preparations of stearoyl-ACP desaturase display one prominent band and several minor bands on SDS–gel electrophoresis. By comparing samples containing various amounts of desaturase and thioesterase, it appears that the prominent band corresponds to the stearoyl-ACP desaturase.

Properties

Specificity. At substrate concentrations of 0.3 μM, the stearoyl-ACP desaturase is 40 times more active on stearoyl-ACP than on stearoyl-CoA and 80 times more active than on palmitoyl-ACP.[8] This high specificity for stearoyl-ACP contrasts with the promiscuous activity of the analogous stearoyl-CoA desaturase from animal systems, which is quite active on acyl-CoA containing 13–19 carbon atoms in the acyl chain.[13]

pH Activity Profile. The desaturase is half-maximally active at pH 5.5 and pH 8.5, with the maximum activity at pH 5.5 in acetate buffer. However, activity at a given pH is dependent on the type of buffer, even at constant ionic strength.[8]

Stability. The stearoyl-ACP desaturase appears to be fairly unstable. It is sensitive to pH, losing 50% or more activity irreversibly when incubated at a pH outside the range pH 6.0 to pH 7.5. It is inactivated on heating at 50° for 1 min. It is unstable to dialysis, irreversibly losing 50% to 100% activity. Finally, further attempts to purify or concentrate the eluent from the ACP-Sepharose column result in nearly total loss of activity.[8]

Miscellaneous Properties. The concentration of oxygen required for maximum activity is 320 μM, which is slightly higher than the oxygen concentration in air-saturated incubation medium, namely 280 μM at 23°; half-maximum activity occurs at a concentration of 60 μM.[8]

Catalase is not required for the desaturase reaction to occur; however, it does stimulate the reaction fivefold. Presumably, catalase protects the desaturase system by scavenging H_2O_2. Both the desaturase and the ferredoxin, NADPH : ferredoxin oxidoreductase system are partially inac-

[13] H. G. Enoch, A. Catala, and P. Strittmatter, *J. Biol. Chem.* **251**, 5095 (1976).

tivated by 0.1 mM H_2O_2, and catalase partially reverses this inactivation.[8] However, two other enzymatic H_2O_2 scavengers do not. Neither horseradish peroxidase nor glutathione peroxidase can replace catalase, and horseradish peroxidase inhibits the desaturation reaction.[8]

[35] Acyl Chain Elongation in Developing Oilseeds

By MICHAEL R. POLLARD

The lipids of most plant tissues contain a narrow spectrum of fatty acids: palmitate, oleate, linoleate, and α-linolenate. The neutral lipids (triacylglycerols)[1] of oilseeds, however, contain a diverse range of fatty acids.[2] One structural variation found is that of acids with chain lengths greater than the usual 16 or 18 carbon atoms. This chapter describes approaches to studying the biosynthesis of acids of chain length C_{20} or greater in developing oilseeds. Some of the considerations noted for the investigation of acyl chain elongation are valid for the investigation of other types of acyl metabolism found in developing oilseeds. Ideally both *in vivo* and *in vitro* experiments are required to demonstrate chain elongation. A radio-gas chromatography machine is useful for detection of ^{14}C-labeled fatty ester.

Supply of Maturing Seed Tissue

The choice of a suitable plant will greatly facilitate the investigation. An ideal plant will exhibit the following features.

1. It should be able to produce a steady supply of developing seeds. That is, a plant is preferred that can be grown and induced to flower all year round, probably in the controlled environment of a growth chamber or greenhouse. A short growth period and early flowering will give maximum experimental flexibility.
2. Larger seeds will help reduce the considerable labor of hand pollination, picking, and removal of the seed coat or pod.
3. For studies on chain elongation, a seed is required that has a high percentage of its fatty acids with a chain length of C_{20} or greater. More than 10% of the dry weight of the mature seed should be lipid

[1] The single exception found in higher plants is the wax esters of jojoba (*Simmondsia chinensis*) seeds [T. W. Miwa, *J. Am. Oil Chem. Soc.* 48, 259 (1971)].

[2] C. Hitchcock and B. W. Nichols, "Plant Lipid Biochemistry," Chapter 1. Academic Press, New York, 1971.

ISBN 0-12-181971-X

in order to measure accurately lipogenic activities *in vivo* and *in vitro*.

Plants that have been used to study the biochemistry of chain elongation in maturing oilseeds are high erucate strains of *Brassica napus* (rape),[3] *Brassica campestris* (turnip rape),[4] and *Brassica juncea* (mustard rape),[5] as well as *Limnanthes alba* (meadowfoam),[6] *Tropaeolum majus* (nasturtium),[7] and *Crambé abyssinica*.[8] They are all annuals. Sometimes, the choice of an oilseed that can be harvested only at a particular time is unavoidable, as in the study of wax ester biosynthesis, which is unique to *Simmondsia chinensis* (jojoba).[9,10] In this case the project becomes distinctly seasonal.

An important preliminary step is to monitor the development of the seeds (Fig. 1). This will ensure that maturing seeds are harvested at the time of maximum lipid biosynthesis. Lipid content (expressed as mass of total or neutral lipid per seed or per gram of fresh or dry seed weight) should be measured as a function of days after flowering (field grown plants) or days after pollination (greenhouse plants). Appelquist has reviewed the topic of lipid accumulation during seed maturation.[11] Several extraction procedures for lipids are suitable. Soxhlet extraction of the dried, ground seeds with petroleum ether will yield neutral lipids.[12] Alternatively, total lipids can be extracted from fresh tissue by homogenizing in petroleum ether–isopropanol, 3 : 2 (v/v)[13] or in chloroform–methanol, 2 : 1 (v/v),[14] followed by the appropriate aqueous salt wash. Extraction of the seed residues should be exhaustive. Developing seeds are ideal for biochemical studies when about 10–50% of the eventual neutral lipid mass has been deposited. Over this period the *in vivo* incorporation of [1-^{14}C]acetate into lipids is generally at a maximum (Fig. 1). Seeds picked later have much endogenous lipid. This can cause severe mass overloading during radio-chromatographic analysis.

[3] R. K. Downey and B. M. Craig, *J. Am. Oil Chem. Soc.* **41,** 475 (1964).
[4] L. A. Appelquist, *J. Am. Oil Chem. Soc.* **50**(2) (1973).
[5] A. Benzioni and M. R. Pollard, unpublished observations, 1979.
[6] M. R. Pollard and P. K. Stumpf, *Plant Physiol.* **66,** 649 (1980).
[7] M. R. Pollard and P. K. Stumpf, *Plant Physiol.* **66,** 641 (1980).
[8] R. S. Appleby, M. I. Gurr, and B. W. Nichols, *Eur. J. Biochem.* **48,** 209 (1974).
[9] J. B. Ohlrogge, M. R. Pollard, and P. K. Stumpf, *Lipids* **13,** 203 (1978).
[10] M. R. Pollard, T. McKeon, L. M. Gupta, and P. K. Stumpf, *Lipids* **14,** 651 (1979).
[11] L. A. Appelquist, *in* "Recent Advances in the Chemistry and Biochemistry of Plant Lipids" (T. Galliard and E. I. Mercer, eds.), pp. 247–286. Academic Press, New York, 1975.
[12] A. Vogel, "A Textbook of Practical Organic Chemistry," 4th ed., p. 137. Longmans Green, New York, 1978.
[13] A. Hara and N. S. Radin, *Anal. Biochem.* **90,** 420 (1978).
[14] J. Folch, M. Lees, and G. H. Sloane-Stanley, *J. Biol. Chem.* **226,** 497 (1957).

ARTICLE

Cloning of Δ12- and Δ6-Desaturases from *Mortierella alpina* and Recombinant Production of γ-Linolenic Acid in *Saccharomyces cerevisiae*

Yung-Sheng Huang[a], Sunita Chaudhary[a], Jennifer M. Thurmond[a], Emil G. Bobik, Jr.[a], Ling Yuan[b,1], George M. Chan[b], Stephen J. Kirchner[a], Pradip Mukerji[a,*], and Deborah S. Knutzon[b]

[a]Ross Products Division, Abbott Laboratories, Columbus, Ohio 43215, and [b]Calgene LLC, Davis, California 95616

ABSTRACT: Two cDNA clones with homology to known desaturase genes were isolated from the fungus *Mortierella alpina*. The open reading frame in one clone encoded 399 amino acids and exhibited Δ12-desaturase activity when expressed in *Saccharomyces cerevisiae* in the presence of endogenous fatty acid substrate oleic acid. The insert in another clone contained an open reading frame encoding 457 amino acids and exhibited Δ6-desaturase activity in *S. cerevisiae* in the presence of exogenous fatty acid substrate linoleic acid. Expression of the Δ12-desaturase gene under appropriate media and temperature conditions led to the production of linoleic acid at levels up to 25% of the total fatty acids in yeast. When linoleic acid was provided as an exogenous substrate to the yeast cultures expressing the Δ6-desaturase activity, the level of γ-linolenic acid reached 10% of the total yeast fatty acids. Co-expression of both the Δ6- and Δ12-desaturase cDNA resulted in the endogenous production of γ-linolenic acid. The yields of γ-linolenic acid reached as high as 8% of total fatty acids in yeast.

Paper no. L8157 in *Lipids 34*, 649–659 (July 1999).

The primary products of fatty acid biosynthesis in most organisms are 16- and 18-carbon compounds. The relative ratio of chain lengths and degree of unsaturation of these fatty acids vary widely among species. Mammals, for example, produce primarily saturated and monounsaturated fatty acids, while most higher plants produce fatty acids with one, two, or three double bonds. Indeed, polyunsaturated fatty acids, such as linoleic acid (Δ9,12-18:2) and α-linolenic acid (Δ9,12,15-18:3), are regarded as essential fatty acids in the diet because mammals lack the ability to synthesize them. However, when ingested, mammals have the ability to metabolize linoleic and α-linolenic acids to form the n-6 and n-3 families of long-chain polyunsaturated fatty acids (LC-PUFA), respectively. These LC-PUFA are important cellular components conferring fluidity to membranes and functioning as precursors of biologically active eicosanoids such as prostaglandins, prostacyclins, and leukotrienes which regulate normal physiological functions (1).

In mammals, the formation of LC-PUFA is rate-limited by the step of Δ6-desaturation, which converts linoleic acid to γ-linolenic acid (GLA, Δ6,9,12-18:3) and α-linolenic acid to stearidonic acid (Δ6,9,12,15-18:4). Many physiological and pathological conditions have been shown to depress this metabolic step, and consequently, the production of LC-PUFA (2). However, bypassing the Δ6-desaturation *via* dietary supplementation with GLA can effectively alleviate many pathological diseases associated with low levels of PUFA (1). This beneficial effect prompted GLA-rich oil to become a much-demanded commodity. GLA is currently used in the treatment of eczema and mastalgia (1). At the present time, the predominant sources of GLA are oils from plants such as borage, evening primrose and black currant, and from microorganisms, such as *Mortierella* spp., *Mucor* spp. and cyanobacteria (3). However, these GLA sources are not ideal for dietary supplementation due to high fluctuations in availability, production/purification costs, unpleasant tastes and odors, and safety concerns. Thus, interest in developing more reliable and economical alternative sources of GLA and other LC-PUFA is growing.

The primary product of fatty acid biosynthesis in most plants and yeast is the monounsaturated, 18-carbon oleic acid. Two desaturation steps, at the Δ12 and Δ6 positions, necessary for the production of GLA from oleic acid, are shown below.

$$\underset{\text{Oleic}}{\Delta 9\text{-}18{:}1} \xrightarrow{\Delta 12\text{-desaturase}} \underset{\text{Linoleic}}{\Delta 9,12\text{-}18{:}2} \xrightarrow{\Delta 6\text{-desaturase}} \underset{\alpha\text{-Linolenic}}{\Delta 6,9,12\text{-}18{:}3}$$

The cDNA clones encoding Δ12-desaturases were isolated from several species of cyanobacteria (4,5) and plants including *Arabidopsis* (6), soybean (7), and parsley (8). Δ6-Desaturase-encoding cDNA were isolated from cyanobacteria (9), borage (10), and nematode (11). These enzymes, as well as numerous examples of Δ15/n-3 desaturases (12,13), are all believed to be integral membrane proteins utilizing an acyl-

[1]Present address: Maxygen, Santa Clara, CA 950510.

*To whom correspondence should be addressed.

E-mail: pradip.mukerji@rossnutrition.com

Abbreviations: GC, gas chromatography; GLA, γ-linolenic acid; LC-PUFA, long-chain polyunsaturated fatty acid; MS, mass spectrometry; PCR, polymerase chain reaction; TPI, triose phosphate isomerase.

lipid substrate, and with the exception of the cyanobacterial enzymes, requiring cytochrome *b5* for the electron transport. The deduced amino acid sequences of these desaturases show a good deal of similarity, most notably in the region of three histidine-rich motifs that are believed to be involved in iron binding (14).

In this study, we utilized the filamentous fungus, *Mortierella alpina*, as the source for desaturase genes. This approach was based on the fact that this fungus is rich in linoleic acid and its LC-PUFA n-6 metabolites, GLA, and arachidonic acid (Δ5,8,11,14-20:4). Using a strategy based on degenerate oligonucleotide primers designed to amplify sequences present at the second and third His boxes of known acyl lipid desaturases (14), we previously isolated a cDNA clone encoding the *M. alpina* Δ5-desaturase (15). A similar strategy utilizing different degenerate primers was also successful in amplifying the same Δ5-desaturase (16). Such polymerase chain reaction (PCR) approaches are limited, however, by the degree of homology of the target cDNA to the particular primers and conditions utilized. In order to achieve a more thorough examination of the fatty acid desaturases present in the fungus, an alternate approach of sequencing random cDNA clones was also employed. Since it was known that the previously characterized membrane-bound Δ12- and Δ15-desaturases, as well as the available cyanobacterial Δ6-desaturase sequences, showed significant amino acid sequence conservation, particularly in the histidine-rich regions, it was postulated that potential *Mortierella* desaturase cDNA could be recognized based on their deduced amino acid sequences. Indeed, this was the strategy that led to the identification of a borage Δ6-desaturase (11) and a castor oleate 12-hydroxylase (17). Because the first histidine-rich motif (His-box) region can occur from 80 to 160 amino acids (240–480 bp) from the N-terminus, and the third region can be roughly 250–300 amino acids (750–900 bp) into the desaturase sequence (14), 300–400 bp of DNA sequence information obtained from the 5′-end of full-length clones might not contain the regions of highest homology among desaturases. Since at the time this work was initiated, no desaturase sequence was identified from *M. alpina* and it was not known how much homology they might display to known sequences, we chose to obtain information from the internal sequences of cDNA clones instead of the 5′-end of full-length clones.

Expression of the *Mortierella* desaturase candidates was carried out in baker's yeast, *Saccharomyces cerevisiae*. This eukaryotic organism was previously shown to be a suitable host containing the necessary cofactors for functional expression of acyl-lipid desaturases. *Saccharomyces cerevisiae* contains a Δ9-desaturase capable of producing monounsaturated palmitoleic and oleic fatty acids, but does not carry out further desaturations. Expression of an *Arabidopsis FAD2* cDNA in *S. cerevisiae* resulted in the production of linoleic and Δ9,12-hexadecadienoic acids from the endogenous oleic acid and palmitoleic acid substrates, respectively (18,19). By culturing *S. cerevisiae* in the presence of exogenous fatty acid substrates, functional expression of a nematode Δ6-desaturase (11) and a fungal Δ5-desaturase (15,16) were demonstrated. In this study, we report the isolation of Δ12- and Δ6-desaturases from *M. alpina*. Simultaneous expression of these two genes in *S. cerevisiae* drives production of GLA at levels of up to 8% of the total fatty acids without the requirement for exogenous fatty acid substrates.

MATERIALS AND METHODS

cDNA library construction. Synthesis of *M. alpina* cDNA was described previously (15). Briefly, double-stranded cDNA were sized fractionated by column chromatography. The two fractions containing the largest cDNA were pooled and packaged to produce a "full-length" library (M7+8) containing *ca.* 6×10^6 clones with an average insert size of 1.77 kb. An additional library, (M11), to be used for random sequencing was constructed by packaging a fraction containing smaller cDNA, which would most likely contain less than full-length clones as well as full-length copies of shorter messages. The average insert size of this library was 1.1 kb; the titer was 240 pfu/μL. Library screening and plaque purification were carried out with the M7+8 library using standard protocols as described previously (20).

Random DNA sequencing. The cDNA-containing plasmids were excised from the λ-ZipLox clones following manufacturer's recommendations (Life Technologies, Gaithersburg, MD). Bacterial cells were plated on ECLB plates containing 50 μg/mL penicillin. DNA sequence was obtained from the 5′-end of the cDNA insert and compared to the National Center for Biotechnology Information nonredundant database using a BLAST server.

Plasmid construction. For expression in yeast, the *M. alpina* cDNA clones for Δ6- and Δ12-desaturase genes were first modified to create *Eco*RI and *Xho*I restriction sites adjacent to the start and stop codons, respectively. Each gene was amplified from the respective cDNA clone using PCR with a pair of primers which have homology to the 5′-end and 3′-end of the gene (restriction sites underlined):

RO-192 (5′-TAGGCTGAATTC**ATG**GCTGCTGCTCCCAGTGTGAGGACG-3′)

and

RO-193 (5′-AACTGCCTCGAG**TTA**CTGCGCCTTACCCATCTTGGAGGC-3′)

are forward and reverse primers with homology to the sequences around the initiation and termination codons of Δ6-desaturase (Ma524), respectively (shown in bold).

RO-194 (5′-TACCTCGAATTC**ATG**GCACCTCCCAACACTATCGATGCC-3″

and

RO-195 (5′-AACCGTCTCGAG**TTA**CTTCTTGAAAAAGACCACGTCTCC-3′)

are forward and reverse primers homologous to the 5′- and 3′-ends of the Δ12-desaturase (Ma648), respectively.

The *Eco*RI/*Xho*I putative desaturase gene fragments were cloned into the vector pYES2 (Invitrogen, San Diego, CA) for inducible expression under the control of GAL1 promoter

in yeast. This vector contains a selectable marker gene which confers uracil prototrophy in the host. The plasmids containing the putative Δ6-desaturase (Ma524) and Δ12-desaturase (Ma648) genes were designated as pCGR-5 and pCGR-7, respectively. To construct pCGR11 and pCGR12, the Δ6- and Δ12-desaturase coding regions were isolated from pCGR5 and pCGR7, respectively, as *Eco*R1-*Xho*I fragments and cloned into the pYX242 vector (Novagen, Madison, WI) digested with *Eco*R1-*Xho*I. The pYX242 vector contains a marker gene for selection of leucine prototrophy in the host and has the promoter of TPI (yeast triose phosphate isomerase gene), which allows constitutive expression. Co-expression of recombinant Δ6- and Δ12-desaturases can be achieved by simultaneous introduction of pCGR5 with pCGR12 or pCGR7 with pCGR11 in the appropriate host requiring both uracil and leucine for growth.

Yeast transformation and expression. Different combinations of pCGR5, pCGR7, pCGR11, and pCGR12 were introduced into a host strain of *S. cerevisiae*, SC334, which contains a mutation (*regl-501*) that alleviates catabolite repression of GAL1 promoter (21). Transformation was done using the PEG/LiAc protocol as described previously (22). Transformants were selected by plating on synthetic medium plates with appropriate selection (21). Cells containing pCGR5 and pCGR7 were selected on media lacking uracil, whereas the pCGR11 and pCGR 12 constructs were selected on media lacking leucine.

Results from our preliminary studies showed that expression of genes (Δ6- and Δ12-desaturases) was enhanced when cultures were grown in synthetic medium at 15°C. In the present study, colonies of transformants were first grown overnight at 30°C in synthetic media. Overnight cultures (2–4 mL) were then used to inoculate 100 mL of minimal media for studying the activities of recombinant desaturases. Galactose was added at a final concentration of 2% to the medium for induction of GAL1 promoter in the strains containing pCGR5 and pCGR7. When the enzyme substrate was provided as the exogenous fatty acid, the fatty acid was supplemented at a concentration of 25 μM. The culture was grown for 48 h at 15°C, and subsequently harvested by centrifugation. Cell pellets were washed once with sterile *dd* H_2O to remove the media. The host strain transformed with vector alone was used as a negative control in all experiments.

Fatty acid analysis. The extraction of the yeast lipids followed the procedures described previously (15). Briefly, washed yeast pellets were extracted with 15 mL of methanol and 30 mL of chloroform containing 100 μg of tridecanoin. After extraction, the yeast lipids were first saponified, and the liberated fatty acids were methylated. The distribution of fatty acid methyl esters was then analyzed by gas chromatography (GC) using a Hewlett-Packard 5890 II Plus gas chromatograph (Hewlett-Packard, Palo Alto, CA) equipped with a flame-ionization detector and a fused-silica capillary column (Supelcomega; 50 m × 0.25 mm, i.d., Supelco, Bellefonte, PA). In the present study, the quantity of the product formed and the rate of conversion of substrate to product (conversion rate = product/(substrate + product)) were calculated to reflect the expression/activity of a given desaturase in this yeast cell assay system.

The identification of a given novel fatty acid was verified by gas chromatography–mass spectrometry (GC–MS) using a Hewlett-Packard mass selective detector (model 5920) operating at an ionization voltage of 70 eV with a scan range of 20–500 Da. The mass spectra of new peaks were compared with those of authentic standards (Nu-Chek-Prep, Elysian, MN) and those in the database NBS75K.L (National Bureau of Standards).

RESULTS

Isolation of a Δ6-desaturase-like cDNA clone from M. alpina. DNA sequence was obtained from the 5′-end of cDNA in randomly picked clones from the *M. alpina* M11 library. Sequence of one such clone, Ma524, exhibited limited homology to a known *Synechocystis* Δ6-desaturase (9) when compared to the databanks. Overall, the level of homology was low (BLAST score 114; P 4.7×10^{-7}). The partial cDNA was used as a probe to isolate a full-length clone, designated pCGN5532, from the M7+8 library. The cDNA insert in pCGN5532 (GenBank accession AF110510) was 1617 bp and contained an open reading frame encoding 457 amino acids flanked by 70 and 75 bp of 5′- and 3′-untranslated regions, respectively. The deduced amino acid sequence is aligned to that of borage Δ6-desaturase (10) in Figure 1. The three "His-boxes," known to be conserved among membrane-bound desaturases (6,14), were found to be present at amino-acid positions 172–176, 209–213, and 395–399 in this sequence. Similar to other membrane-bound Δ6- and Δ5-desaturases, the final "HXXHH" histidine box motif was found to be QXXHH (11,15,16). The predicted amino-acid sequence from this clone is similar to the Δ6-desaturases from the *Synechocystis* spp. and *Spirulina* spp. (9), the borage Δ6-desaturase (10), the nematode *Caenorhabditis elegans* (11), and a cytochrome *b5*/desaturase fusion protein from sunflower (23). As reported for other Δ5/Δ6 desaturases, the amino terminus of the protein encoded by pCGN5532 was also homologous to cytochrome *b5* proteins.

Isolation of a Δ12-desaturase-like cDNA clone from M. alpina. DNA sequence obtained from the 5′-end of another random clone, Ma648, showed homology to the soybean n-6 desaturase (7). The homology of the partial *M. alpina* sequence was again relatively weak (BLAST score 110, P 2.0×10^{-6}). Analysis of the open reading frames beginning at the 5′-end of Ma648 indicated that the first possible methionine was in frame +1 which was the frame that showed desaturase homology. Alignment of this open reading frame to 5′-sequence of other Δ12-desaturases indicated that the *M. alpina* Ma648 clone was full-length. This cDNA was designated pCGN5533, and no other corresponding clones were obtained by library screening. The 1488 bp cDNA insert in pCGN5533 (GenBank accession AF110509) contains 78 bp of 5′- and 113 bp of 3′-noncoding sequences flanking an open

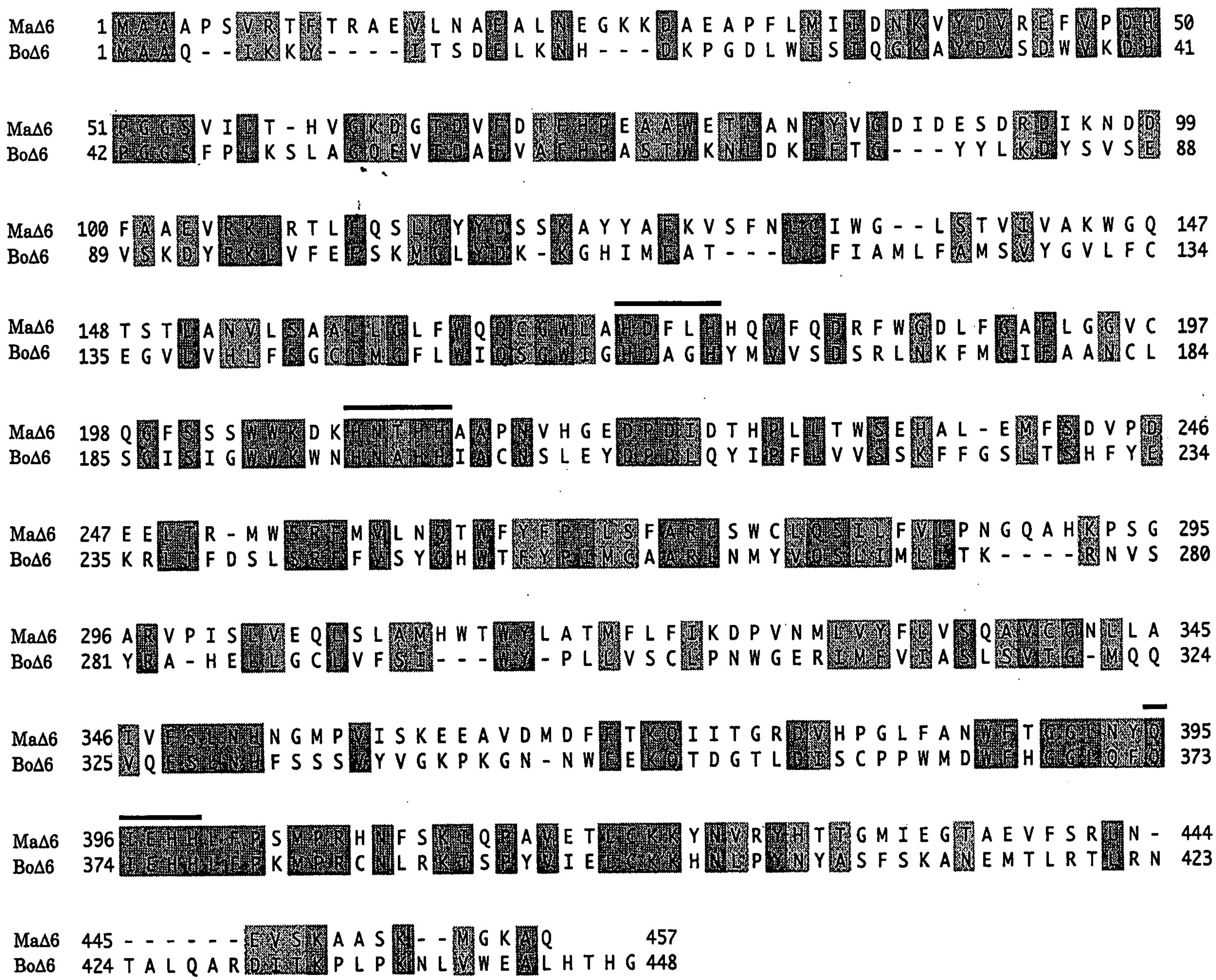


FIG. 1. Comparison of the deduced amino-acid sequences of *Mortierella alpina* and a prototype Δ6-desaturase from borage. Conserved amino acids are shaded light; identical residues are shaded dark. Ma Δ6, *M. alpina* Δ6-desaturase, pCGN5532; Bo 6, Δ6-desaturase from *Borago officinalis* (10). The three regions containing histidine residues (His-boxes) conserved among acyl-lipid desaturases and hydroxylases are overlined.

reading frame encoding 399 amino acids. Figure 2 shows the alignment of the deduced amino acid sequence of pCGN5533 to the *FAD2* (microsomal Δ12 desaturase) from *Arabidopsis* (6). The three His-boxes are again present at positions 111–115, 147–151, and 338–342. Unlike Ma524, no homology to cytochrome *b5* sequence is present on the N-terminus of this clone.

Functional expression of M. alpina *desaturase clone pCGN5533 (Ma648) in yeast.* In order to assess the functional specificity of the various *M. alpina* desaturase clones, the coding regions were expressed in *S. cerevisiae* using the inducible GAL1 promoter found in the commercial vector pYES2. As described previously (15), recombinant yeast cells were grown in the presence of various fatty acids in order to provide substrates for desaturases involved in LC-PUFA production. The deduced coding region of pCGN5533 (Ma648) was inserted into the yeast expression vector pYES2 to create pCGR7. Fatty acid profiles of lipid fractions from yeast grown in the absence of exogenous fatty acid substrate show that two novel fatty acids were produced in SC334(pCGR7) (Fig. 3A). The first fatty acid showed a mass peak $m/z = 266$ (the expected molecular ion of 16:2), a retention time of 13.48 min, and a fragmentation pattern identical to those of Δ9,12-16:2 (Fig. 3B). The second novel fatty acid exhibited a retention time (17.28 min) in GC (Fig. 3A), mass peak (m/z = 294) and fragmentation pattern in GC–MS (data not shown) identical to those of the authentic linoleic acid (Δ9,12-18:2). These findings indicate that the endogenous oleic acid (Δ9-18:1) was converted to linoleic acid (Δ9,12-18:2) by a Δ12-desaturase activity expressed from the plasmid pCGR7. The rate of conversion was found to be 71.4% (Table 1).

Functional expression of M. alpina *desaturase clone pCGN5532 (Ma524) in yeast.* The recombinant yeast SC334(pCGR5), containing the Ma524 cDNA, was grown in the presence of exogenous linoleic acid (Δ9,12-18:2) which is the substrate for Δ6-desaturation. Analyses of the fatty acid profile in the yeast lipid fraction indicate that the exogenous linoleic acid was incorporated into lipids of both nontransformed and transformed yeast (Fig. 4). However, GC analysis revealed the presence of a novel fatty acid in the SC334(pCGR5) yeast that was not present in yeast transformed with vector alone. This novel fatty acid had a retention time of 17.96 min in GC (Fig. 4). Mass peak $m/z = 292$ and fragmentation pattern of this fatty acid in GC–MS were identical to those of the authentic GLA (Δ6,9,12-18:3) standard; however, the fragmentation pattern was different from that of the α-linolenic acid (Δ9,12,15-18:3) standard (data not shown). Thus, the Ma524 cDNA expressed from pCGR5 encodes a Δ6-desaturase. The expressed enzyme converted 29.4% of the incorporated linoleic acid to GLA (Table 1).

Since there were no traces of α-linolenic acid (Δ9,12,15-18:3) produced from the exogenous linoleic acid (Δ9,12-18:2) in the recombinant yeast strains, it is suggested that the enzyme produced by pCGR5 does not possess Δ15-desaturase activity. In addition, when exogenous α-linolenic acid was included in the growth medium, 3.9% of the incorporated α-linolenic acid (Δ9,12,15-18:3) was converted to stearidonic acid (Δ6,9,12,15-18:4). The identity of stearidonic acid was verified by both GC and GC–MS (data not shown). This finding further confirms the enzyme to be a Δ6-desaturase.

TABLE 1
Production of Linoleic Acid and GLA in Yeast Lipid Fraction

SC334 containing	Total fatty acids[a] (μg)	Oleic (wt%)	Linoleic (wt%)	GLA[c] (wt%)
pYES2	440.1	23.2	—	—
pCGR5[b]	497.1	10.2	25.1	10.3
pCGR7	460.9	10.0	24.8	—
pCGR11/pCGR7	340.8	10.2	10.1	7.9
pCGR5/pCGR12	367.9	6.7	7.0	6.6

[a]The volume of culture used for lipid extraction was 100 mL.
[b]Exogenous linoleic acid (25 μM) was added.
[c]No α-linolenic acid was detected. GLA, γ-linolenic acid.

In the absence of exogenous linoleic acid, the lipid fraction of the yeast strain expressing the Δ6-desaturase cDNA produced two novel fatty acids (Fig. 5A). The first novel fatty acid showed a mass peak $m/z = 266$, which is the expected molecular ion of 16:2. Although the GC–MS fragmentation patterns of this novel fatty acid and the authentic Δ9,12-16:2 were similar, they were different in intensity (Figs. 3B and 5B), and retention time (12.89 vs. 13.48 min) in GC (Figs. 3A and 5A). Since this novel fatty acid was produced in the presence of the Δ6-desaturase, it was most probably the Δ6,9-16:2. The second novel fatty acid produced in SC334(pCGR5) had an identical retention time (16.95 min) in GC (Fig. 5A), mass peak $m/z = 294$, and fragmentation pattern in GC–MS to that of the Δ6,9-18:2 standard (data not shown).

Production of GLA. As shown above, the recombinant Δ12- and Δ6-desaturases were effective in converting their substrates (endogenous oleic acid and exogenous linoleic acid) to their respective products, linoleic acid (Fig. 3) in SC334(pCGR7) and GLA (Fig. 4) in SC334(pCGR5). We were interested in determining the feasibility of producing GLA in a recombinant yeast strain in the absence of exogenously added fatty acid substrates. The biosynthesis of GLA from the endogenous oleic acid in *S. cerevisiae* would require the simultaneous expression of Δ12- and Δ6-desaturases. In order to allow co-expression of the Δ6- and Δ12-desaturase cDNA, they were cloned under the control of the constitutive TPI promoter into the leucine-selectable vector pYX242 to create pCGR11 and pCGR12. Both combinations of promoters GAL1 and TPI were assayed for production of GLA. The co-expression of pCGR11 (containing Δ6-desaturase gene under the control of TPI) and pCGR7 (containing Δ12-desaturase gene under the control of GAL1) resulted in *ca.* 7.9% of GLA in total fatty acids of SC334(pCGR7, pCGR11) (Table 1). The rates of conversion from oleic acid to linoleic acid and from linoleic acid to GLA were *ca.* 50 and 44%, respectively. In SC334 (pCGR12, pCGR5) containing the Δ6-desaturase gene behind GAL1 and the Δ12-desaturase gene behind TP1, a level of 6.6% of GLA was found in the total fatty acids (Table 1). In these recombinant yeast strains, the

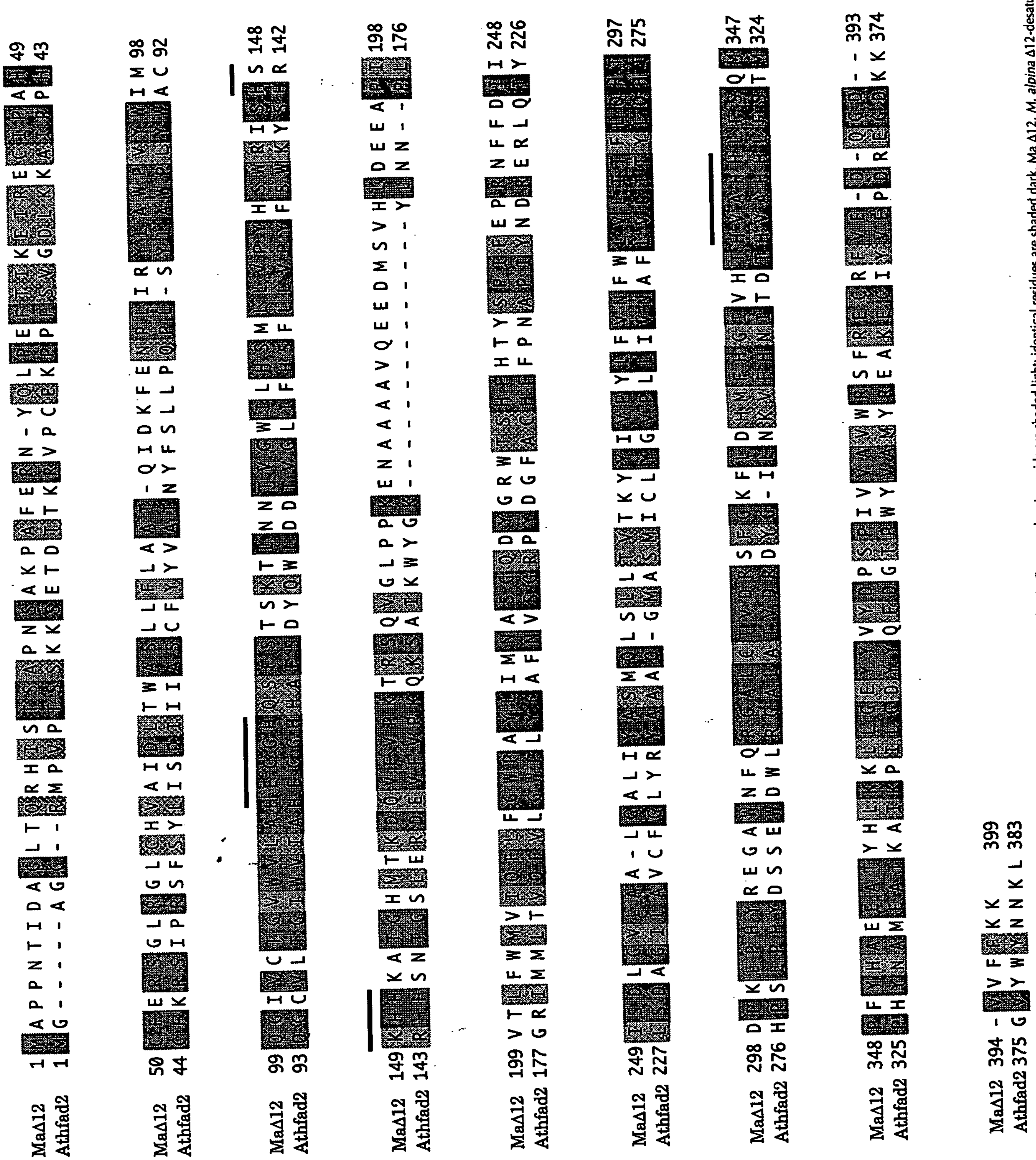


FIG. 2. Comparison of the deduced amino-acid sequences of *M. alpina* and a prototype Δ12-desaturase from *Arabidopsis*. Conserved amino acids are shaded light; identical residues are shaded dark. Ma Δ12, *M. alpina* Δ12-desaturase, pCGN5533; Athfad2, Δ12-desaturase from *A. thaliana* (6). The three regions containing histidine residues (His-boxes) conserved among acyl-lipid desaturases and hydroxylases are overlined. See Figure 1 for abbreviation.

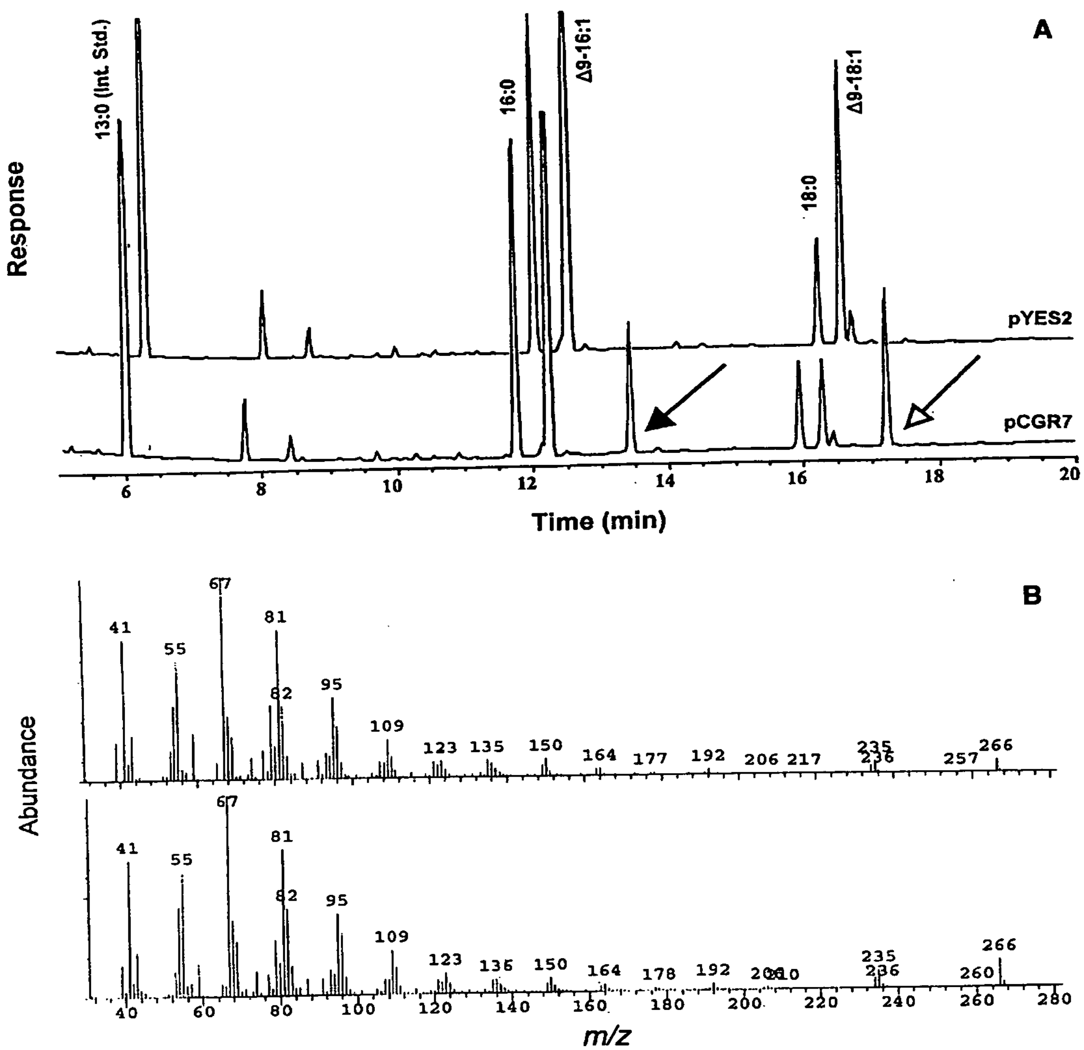


FIG. 3. (A) Gas chromatographic analysis of fatty acid methyl esters (FAME) from the lipid fraction in yeast containing pYES2 or pCGR7. Solid and open arrows indicate the fatty acids Δ9,12-16:2 and Δ9,12-18:2, respectively, present in SC334(pCGR7) cultures. (B) Gas chromatography–mass spectrometry (GC–MS) analysis of the novel peak (identified by the solid arrow in Fig. 3A) in yeast carrying pCGR7. The fragmentation pattern of the first novel peak (top) was compared with that of the authentic Δ9,12-16:2 standard (bottom). pYES2 contained only vector whereas pCGR7 contained the coding region of the *M. alpina* Δ12-desaturase cDNA clone, pCGN5533. All yeast strains were grown in the minimal medium. See Figure 1 for other abbreviations.

conversion rate for both oleic acid to linoleic acid and linoleic acid to GLA was about 50%. Among them, SC334(pCGR11, pCGR7) produced a higher level of GLA, and the GLA accumulated predominantly in the phospholipid fraction (data not shown). Hence, co-expression of *M. alpina* Δ6- and Δ12-desaturase genes under the control of independent promoters in yeast resulted in *de novo* synthesis of GLA.

Comparison of desaturase amino acid sequences. The availability of three desaturase sequences from *M. alpina* was used to examine the interspecies and interclass relationships of these sequences. The amino-acid sequences between the first and third His-boxes of representative desaturases were used to construct a similarity dendrogram (Fig. 6). Two major classes of desaturases can be discerned in this dendrogram. One class contains the Δ12/n-6 and Δ15/n-3 desaturases, while all known examples of Δ5- or Δ6-desaturases fall into a separate class. Although only the central amino-acid core sequence was used in the alignments, all the desaturase sequences with an N-terminal cytochrome *b5*-like sequence cluster into the latter sequence group. The presence of the cytochrome *b5* extension appears to be related to the functionality of the desaturase and not the source of the gene; of the three desaturases from *M. alpina*, only the Δ5- and Δ6-desaturases have the fused cytochrome sequence. In addition, all of the members of this class contain the H-Q substitution in the third His box.

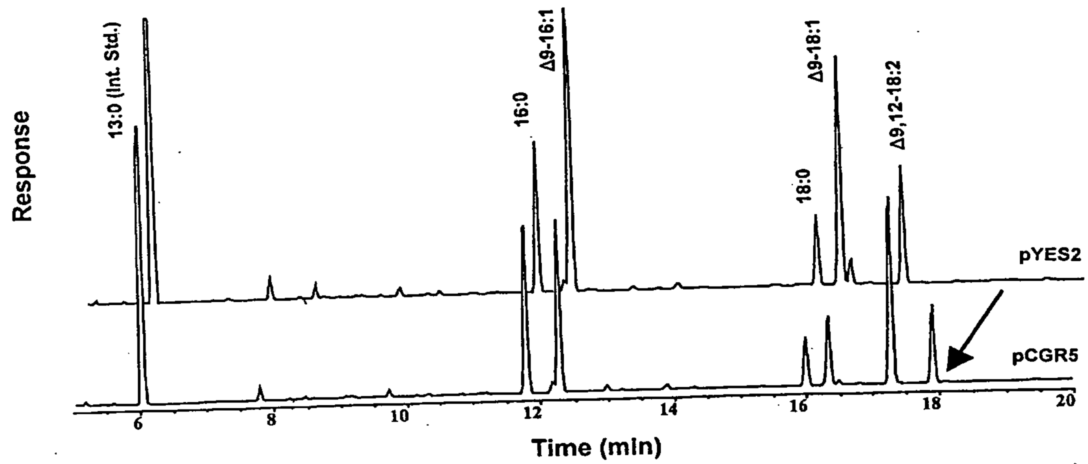


FIG. 4. Gas chromatographic analysis of FAME from the lipid fraction in yeast containing pYES2 or pCGR5. The arrow indicates a novel fatty acid (Δ6,9,12-18:3) present in SC334(pCGR5) cultures. pYES2 contained only vector, whereas pCGR5 contained the coding region of the *M. alpina* Δ6-desaturase cDNA clone, pCGN5532. All yeast strains were grown in minimal medium supplemented with the exogenous linoleic acid (Δ9,12-18:2). See Figures 1 and 3 for abbreviations.

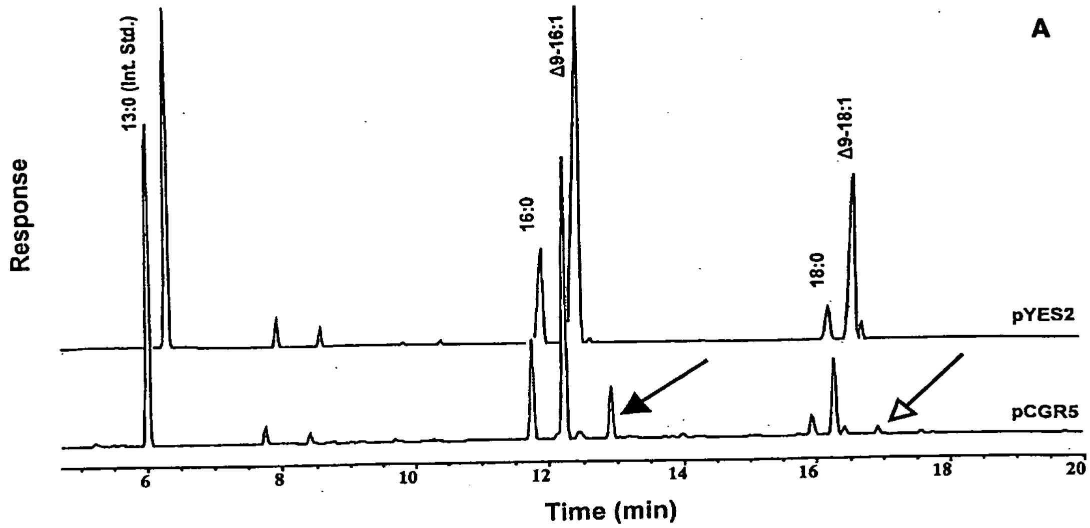
DISCUSSION

We utilized a random sequencing approach to identify cDNA clones encoding fatty acid desaturases from the fungus, *M. alpina*. Partial sequence obtained from the 5′-end of randomly-selected clones was compared to the databanks, and homologies to known acyl-lipid desaturases were noted. This report describes the isolation of two different desaturase-like cDNA clones encoding Δ6- and Δ12-desaturases. These clones were identified in a first-phase sequencing of ~1200 cDNA. In addition to the two clones described in this work, the first phase of sequencing also revealed clones corresponding to the Δ5-desaturase originally obtained by heterologous PCR (15,16) and clones homologous to the yeast stearoyl-CoA desaturase (24) (data not shown). A more thorough sequencing effort of 5400 additional cDNA resulted in the identification of 13 sequences with homology to stearoyl-CoA desaturases, 8 Δ6-desaturases, 9 Δ5-desaturases, and 5 Δ12-desaturases. It should be noted that several of the random clones encoding Δ5- and Δ6-desaturases actually showed cytochrome *b5* matches in the BLAST results, due to the highly homologous cytochrome domain at the N-terminus of these desaturases. Had this domain not been previously identified, several of these cDNA might not have been recognized as desaturases in such a mass sequencing effort. This is an important point to keep in mind when interpreting BLAST results of all sequences; the presence of one highly conserved domain may lead to mis-annotation of the sequence.

The comparison of the desaturase amino-acid sequences shown in Figure 6 indicates that the *M. alpina* Δ5-desaturase is more closely related to the cyanobacterial Δ6-desaturases than to the plant and animal Δ6-desaturase sequences. The ultimate significance of this is hard to evaluate, due to the lack of other Δ5-desaturases for comparison. It should, however, be noted that the *C. elegans* ORF on cosmid T13F2 (GenBank accession number Z81122) that was proposed to be a possible Δ5-desaturase (16) shows more similarity to the *M. alpina* Δ6-desaturase sequence than to the *M. alpina* Δ5-desaturase sequence (data not shown).

In the present study, we showed that the recombinant enzyme expressed by a *M. alpina* desaturase-like gene (Ma648) in pCGR7 converted Δ9-16:1 to Δ9,12-16:2 and oleic acid (Δ9-18:1) to linoleic acid (Δ9,12-18:2) (Fig. 3A). These findings clearly demonstrated that this gene encodes the Δ12-desaturase. We also showed that the recombinant enzyme expressed by another *M. alpina* gene (Ma524) in pCGR5 converted n-6 fatty acid linoleic acid (Δ9,12-18:2) to GLA (Δ6,9,12-18:3) (Fig. 4). When an n-3 fatty acid, α-linolenic acid (Δ9,12,15-18:3), was used as the substrate, the SC334(pCGR5) produced the expected product, stearidonic acid (Δ6,9,12,15-18:4) (data not shown). In the absence of linoleic acid as substrate, this recombinant enzyme could convert the endogenous Δ9-16:1 to Δ6,9-16:2, and oleic acid (Δ9-18:1) to Δ6,9-18:2 (Fig. 5A). These findings demonstrate that this gene encodes the Δ6-desaturase.

In order to evaluate the feasibility of producing GLA—a high-value PUFA in this microorganism, we co-expressed the genes encoding Δ6- and Δ12-desaturases in yeast. When both genes were presented in a single construct in yeast, and expressed from a single promoter, GAL1, none of these transformed yeast strains produced a significant amount of GLA (data not shown). Therefore, it seemed likely that two independent promoters would be needed for the concurrent expression of these two desaturases. Indeed, when these desaturases were co-expressed *in trans* from two independent promoters, GAL1 and TPI, the production of GLA reached as high as 8% of the total lipids in yeast grown without exogenous substrates (Table 1). By the action of two separate pro-



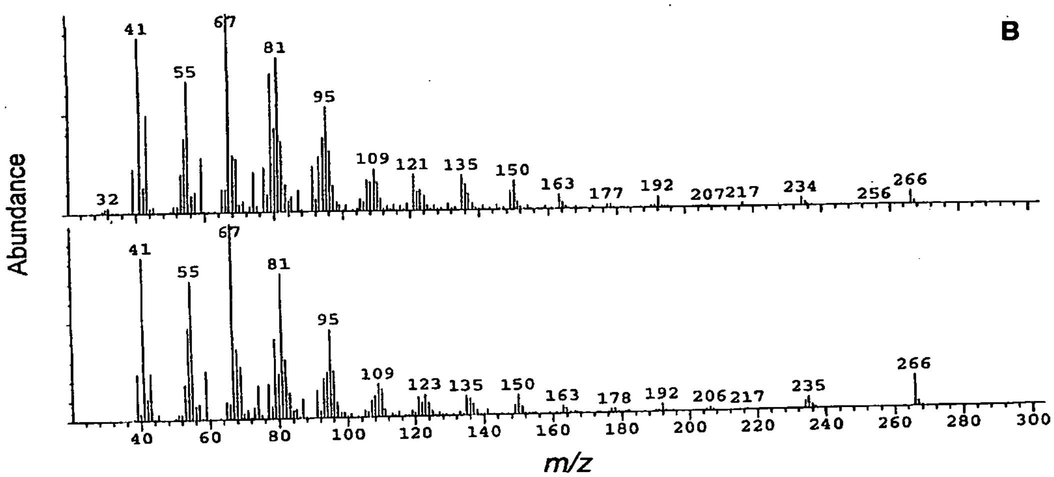


FIG. 5. (A) Gas chromatographic analysis of FAME from the lipid fraction in yeast containing pYES2 or pCGR5 grown without exogenous substrate. Solid and open arrows indicate novel fatty acids Δ6,9-16:2 and Δ6,9-18:2, respectively. (B) GC–MS analysis of the novel peak (identified by the solid arrow in panel A) in yeast carrying pCGR5. The fragmentation pattern of the first novel peak was compared with that of the authentic Δ6,9-16:2 standard pYES2 contained only vector, whereas pCGR5 contained *M. alpina* cDNA clone encoded with Δ6-desaturase. See Figures 1 and 3 for abbreviations.

moters, these enzymes were able to effectively convert (*ca.* 50%) their respective substrates to products.

In summary, we isolated two cDNA from *M. alpina* encoding the Δ6- and Δ12-desaturase genes using a random sequencing-based strategy. The identities of the two cDNA confirmed by functional expression and analysis in a widely used microorganism, baker's yeast. By introducing the two required desaturases (Δ6- and Δ12-) under the control of independent promoters in yeast, we developed a novel approach to synthesize GLA.

ACKNOWLEDGMENTS

The nucleotide sequences reported in this paper were submitted to the GenBank/EBI Data Bank with accession numbers AF110509 and AF110510.

REFERENCES

1. Horrobin, D.F. (1992) Nutritional and Medical Importance of Gamma-Linolenic Acid, *Prog. Lipid Res. 31*, 163–194.
2. Brenner, R.R. (1976) Regulatory Function of Δ6-Desaturase—A Key Enzyme of Polyunsaturated Fatty Acid Synthesis, *Adv. Exp. Med. Biol. 83*, 85–101.
3. Phillips, J.C., and Huang, Y.-S. (1996) Natural Sources and Biosynthesis of γ-Linolenic Acid: An Overview, in *γ-Linolenic Acid: Metabolism and Its Roles in Nutrition and Medicine* (Huang, Y.-S., and Milles, D.E., eds.), AOCS Press, Champaign, pp. 1–13.
4. Sakamoto, T., Wada, H., Nishida, I., Ohmori, M., and Murata,

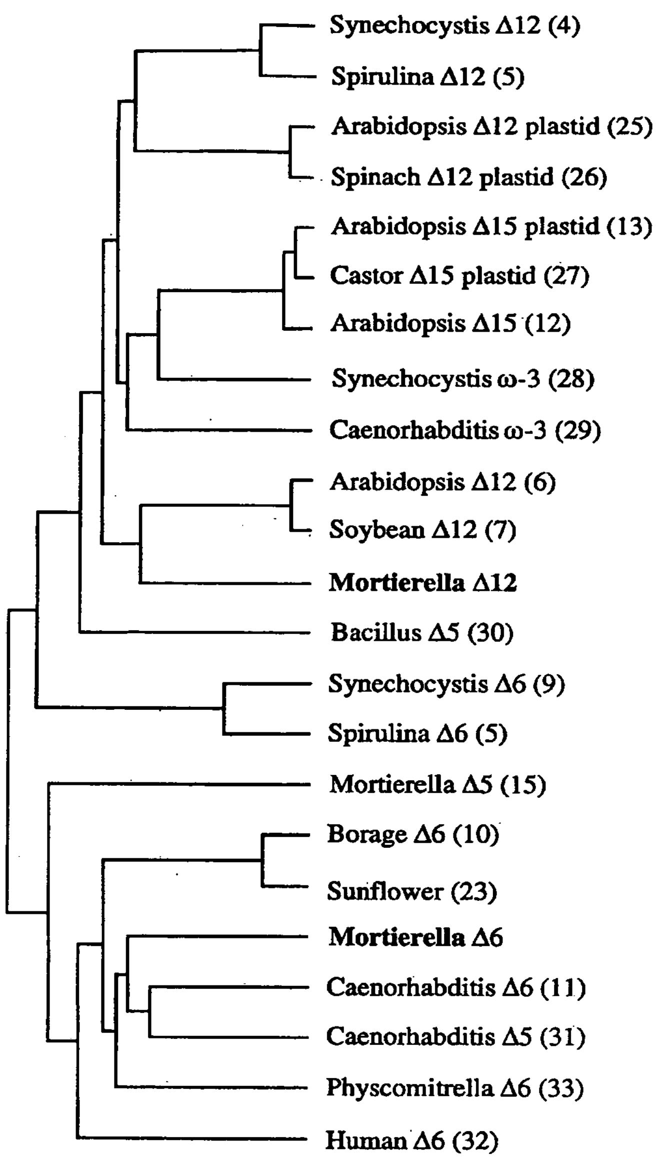


FIG. 6. Similarities of representative membrane-bound desaturases. Dendrogram was constructed using the CLUSTAL program to align deduced amino-acid sequences between the first and third His boxes. Numbers in parentheses indicate the references for the sequences; *Mortierella* Δ6- and Δ12-desaturase sequences are described in this work.

N. (1994) Identification of Conserved Domains in the Δ12 Desaturases of Cyanobacteria, *Plant Mol. Biol. 24*, 643–650.

5. Murata, N., Deshnium, P., and Tasaka, Y. (1996) Biosynthesis of Gamma-Linolenic Acid in the Cyanobacterium *Spirulina platensis*, in *γ-Linolenic Acid: Metabolism and Its Roles in Nutrition and Medicine* (Huang, Y.-S., and Milles, D.E., eds.), AOCS Press, Champaign, pp. 22–32.
6. Okuley, J., Lightner, J., Feldmann, K., Yadav, N., Lark, E., and Browse, J. (1994) *Arabidopsis FAD2* Gene Encodes the Enzyme That Is Essential for Polyunsaturated Lipid Synthesis, *Plant Cell 6*, 147–158.
7. Heppard, E.P., Kinney, A.J., Stecca, K.L., and Miao, G.-H. (1996) Developmental and Growth Temperature Regulation of Two Different Microsomal ω-6 Desaturase Genes in Soybeans, *Plant Physiol. 110*, 311–319.
8. Kirsch, C., Hahlbrock, K., and Somssich, I.E. (1997) Rapid and Transient Induction of a Parsley Microsomal Δ12 Fatty Acid Desaturase mRNA by Fungal Elicitor, *Plant Physiol. 115*, 283–289.
9. Reddy, A.S., Nuccio, M.L., Gross, L.M., and Thomas, T.L. (1993) Isolation of a Δ^6-Desaturase Gene from the Cyanobacterium *Synechocystis* sp. Strain PCC6803 by Gain-of-Function Expression in *Anabaena* sp. Strain PCC7120, *Plant Mol. Biol. 27*, 293–300.
10. Sayanova, O., Smith, M.A., Lapinsk, P., Stobart, A.K., Dobson, G., Christie, W.W., Shewry, P.R., and Napier, J.A. (1997) Expression of a Borage Desaturase cDNA Containing an N-Terminal Cytochrome *b5* Domain Results in the Accumulation of High Levels of Δ^6-Desaturated Fatty Acids in Transgenic Tobacco, *Proc. Natl. Acad. Sci. USA 94*, 4211–4216.
11. Napier, J.A., Hey, S.J., Lacey, D.J., and Shewry, P.R. (1998) Identification of a *Caenorhabditis elegans* Δ^6-Fatty-Acid-Desaturase by Heterologous Expression in *Saccharomyces cerevisiae*, *Biochem. J. 330*, 611–614.
12. Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H., and Somerville, C.R. (1992) Map-Based Cloning of a Gene Controlling Omega-3 Fatty Acid Desaturation in *Arabidoposis*, *Science 258*, 1353–1355.
13. Yadav, N.S., Wierzbicki, A., Aegerter, M., Caster, C.S., Perez-Grau, L., Kinney, A.J., Hitz, W.D., Booth, R., Schweiger, B., Stecca, K.L., Allen, S.M., Blackwell, M., Reiter, R.S., Carlson, T.J., Russell, S.H., Feldmann, K.A., Pierce, J., and Browse, J. (1993) Cloning of Higher Plant ω-3 Fatty Acid Desaturases, *Plant Physiol. 103*, 467–476.
14. Shanklin, J., White, E., and Fox, B.G. (1994) Eight Histidine Residues Are Catalytically Essential in a Membrane-Associated Iron Enzyme, Stearoyl-CoA Desaturase, and Are Conserved in Alkane Hydroxylase and Xylene Monooxygenase, *Biochemistry 33*, 12787–12794.
15. Knutzon, D.S., Thurmond, J.A., Huang, Y.-S., Chaudhary, S., Bobik, E.G., Jr., Chan, G.M., Kirchner, S.J., and Mukerji, P. (1998) Identification of Δ5-Desaturase from *Mortierella alpina* by Heterologous Expression in Baker's Yeast and Canola, *J. Biol. Chem. 273*, 29360–29366.
16. Michaelson, L.V., Lazarus, C.M., Griffiths, G., Napier, J.A., and Stobart, A.K. (1998) Isolation of a Δ5-Fatty Acid Desaturase Gene from *Mortierella alpina*, *J. Biol. Chem. 273*, 19055–19059.
17. van de Loo, F.N., Broun, P., Turner, S., and Somerville, C.R. (1995) An Oleate 12-Hydroxylase from *Ricinus communis* L. Is a Fatty Acyl Desaturase Homolog, *Proc. Natl. Acad. Sci. USA 92*, 6743–6747.
18. Covello, P.S., and Reed, D.W. (1996) Functional Expression of the Extraplastidial *Arabidopsis thaliana* Oleate Desaturase Gene (*FAD2*) in *Saccharomyces cerevisiae*, *Plant Physiol. 111*, 223–226.
19. Kajiwara, S., Shirai, A., Fujii, T., Toguri, T., Nakamura, K., and Ohtaguchi, K. (1996) Polyunsaturated Fatty Acid Biosynthesis in *Saccharamyces cerevisiae*: Expression of Ethanol Tolerance and the *FAD2* Gene from *Arabidopsis thaliana*, *Appl. Env. Microbiol. 62*, 4309–4313.
20. Knutzon, D.S., Lardizabal, K.D., Nelsen, J.S., Bleibaum, J.L., Davies, H.M., and Metz, J.G. (1995) Cloning of a Coconut Endosperm cDNA Encoding a 1-Acyl-*sn*-glycerol-3-phosphate Acyltransferase That Accepts Medium-Chain-Length Substrates, *Plant Physiol. 109*, 999–1006.
21. Hoveland, P., Flick, J., Johnston, M., and Sclafani, R.A. (1989) Galactose as a Gratuitous Inducer of GAL Gene Expression in Yeasts Growing on Glucose, *Gene 83*, 57–64.

22. Sterns, T., Ma, H., and Botstein, D. (1990) Manipulating Yeast Genome Using Plamid Vectors, *Meth. Enzymol. 185*, 280–297.
23. Sperling, P., Schmidt, H., and Heinz, E. (1995) A Cytochrome-*b5*-Containing Fusion Protein Similar to Plant Acyl Lipid Desaturases, *Eur. J. Biochem. 232*, 798–805.
24. Stukey, J.E., McDonough, V.M., and Martin, C.E. (1990) The *OLE1* Gene of *Saccharomyces cerevisiae* Encodes the Δ9-Fatty Acid Desaturase and Can Be Functionally Replaced by the Rat Stearoyl-CoA Desaturase Gene, *J. Biol. Chem. 265*, 20144–20149.
25. Falcone, D.L., Gibson, S., Lemieux, B., and Somerville, C. (1994) Identification of a Gene That Complements an *Arabidopsis* Mutant Deficient in Chloroplast ω6 Desaturase Activity, *Plant Physiol. 106*, 1453–1459.
26. Schmidt, H., Dresselhaus, T., Buck, F., and Heinz, E. (1994) Purification and PCR-Based cDNA Cloning of a Plastidial n-6 Desaturase, *Plant Mol. Biol. 26*, 631–642.
27. van de Loo, F.J., and Somerville, C. (1994) Plastid ω-3 Fatty Acid Desaturase cDNA from *Ricinus communis*, *Plant Physiol. 105*, 443–444.
28. Sakamoto, T., Los, D.A., Higashi, S., Wada, H., Nishida, I., Ohmori, M., and Murata, N. (1994) Cloning of ω3 Desaturase from Cyanobacteria and Its Use in Altering the Degree of Membrane-Lipid Unsaturation, *Plant Mol. Biol. 26*, 249–263.
29. Spychalla, J.P., Kinney, A.J., and Browse, J. (1997) Identification of an Animal ω-3 Fatty Acid Desaturase by Heterologous Expression in *Arabidopsis*, *Proc. Natl. Acad. Sci. USA 94*, 1142–1147.
30. Aguilar, P.S., Cronan, J.E., Jr., and de Mendoza, D. (1998) A *Bacillus subtilis* Gene Induced by Cold Shock Encodes a Membrane Phospholipid Desaturase, *J. Bacteriol. 180*, 2194–2200.
31. Watts, J.L., and Browse, J. (1999) Isolation and Characterization of a Δ^5-Fatty Acid Desaturase from *Caenorhabditis elegans*, *Arch. Biochem. Biophys. 362*, 175–182.
32. Cho, H.P., Nakamura, M.T., and Clarke, S.D. (1999) Cloning, Expression, and Nutritional Regulation of the Mammalian Δ-6 Desaturase, *J. Biol. Chem. 274*, 471–477.
33. Girke, T., Schimdt, H., Zahringer, U., Reski, R., and Heinz, E. (1998) Identification of a Novel Δ6-Acyl-Group Desaturase by Targeted Gene Disruption in *Physcomitrella patens*, *Plant J. 15*, 39–48.

[Received February 17, 1999, in final revised form May 28, 1999; revision accepted June 3, 1999]

nature

INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Volume 347 No.6288 6 September 1990 £1.95

MAKING SMOOTHER WAVES

Actin structure solved

Ozone loss in the Arctic

SCIENTIFIC COMPUTING product review

cells, although the number of endogenous *XMyoD* transcripts is much less than that normally found in the myotomes (Table 1).

Do animal caps isolated from *XMyoD*-injected embryos show other signs of muscle differentiation? The *XMyoD*-injected animal caps are histologically indistinguishable from uninjected or *XMyoD114P*-injected controls (Fig. 3*a, c, d*). By contrast, uninjected animal caps induced by vegetal tissue elongate and contain large blocks of muscle (Fig. 3*b*). We obtained intense labelling of this muscle tissue using the 12/101 anti-muscle antibody[21], but saw no labelling above background of the *XMyoD*-injected animal cap cells (Fig. 3). We conclude that no differentiated muscle is formed by *XMyoD*-injected animal caps. Thus, animal cap cells that contain as much cardiac actin RNA as normal myotomal cells do not express the full myogenic programme, but continue to differentiate as epidermis. Whole *XMyoD*-injected embryos also develop relatively normally, becoming tadpoles with substantially normal external and internal structures, including a variety of differentiated cell types (data not shown).

We have shown that XMyoD can activate a muscle gene to its normal level in animal cap cells. MyoD can bind to sites in the promoters of muscle genes[2], and we note that, in addition to a CArG sequence that is essential for transcription[22], there are potential XMyoD-binding sites[2] located further upstream in the *Xenopus* cardiac actin promoter (M. V. Taylor, N.D.H. and T. J. Mohun, unpublished data). The lack of muscle differentiation in *XMyoD*-injected animal caps could be caused by a failure to maintain a sufficiently high concentration of XMyoD protein. Alternatively, as enough XMyoD has been supplied to activate the cardiac actin gene to its normal level, it may be that muscle did not differentiate because other myogenic factors, not themselves activated by XMyoD, are required to divert these embryonic cells from their normal pathway of differentiation.

Received 20 April; accepted 4 July 1990.

1. Davis, R. L., Weintraub, H. & Lassar, A. B. *Cell* **51**, 987–1000 (1987).
2. Lassar, A. B. *et al. Cell* **58**, 823–831 (1989).
3. Weintraub, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 5434–5438 (1989).
4. Schäfer, B. W., Blakely, B. T., Darlington, G. J. & Blau, H. M. *Nature* **344**, 454–458 (1990).
5. Hopwood, N. D., Pluck, A. & Gurdon, J. B. *EMBO J.* **8**, 3409–3417 (1989).
6. Sassoon, D. *et al. Nature* **341**, 303–307 (1989).
7. Harvey, R. P. & Melton, D. A. *Cell* **53**, 687–697 (1988).
8. Kintner, C. R. *Neuron* **1**, 545–555 (1988).
9. Ruiz i Altaba, A. & Melton, D. A. *Cell* **57**, 317–326 (1989).
10. McMahon, A. P. & Moon, R. T. *Cell* **58**, 1075–1084 (1989).
11. Krieg, P. A. & Melton, D. A. *Nucleic Acids Res.* **12**, 7057–7071 (1984).
12. Murre, C., McCaw, P. S. & Baltimore, D. *Cell* **56**, 777–783 (1989).
13. Davis, R. L., Cheng, P.-F., Lassar, A. B. & Weintraub, H. *Cell* **60**, 733–746 (1990).
14. Nieuwkoop, P. D. *Wilhelm Roux Arch. dev. Biol.* **162**, 341–373 (1969).
15. Sudarwati, S. & Nieuwkoop, P. D. *Wilhelm Roux Arch. dev. Biol.* **169**, 189–204 (1971).
16. Nieuwkoop, P. D. & Faber, J. *Normal Table of* Xenopus laevis *(Daudin)* (North-Holland, Amsterdam, 1967).
17. Mohun, T. J., Brennan, S., Dathan, N., Fairman, S. & Gurdon, J. B. *Nature* **311**, 716–721 (1984).
18. Thayer, M. J. *et al. Cell* **58**, 241–248 (1989).
19. Braun, T. *et al. EMBO J.* **8**, 3617–3625 (1989).
20. Lin, Z., Dechesne, C. A., Eldridge, J. & Paterson, B. M. *Genes Dev.* **3**, 986–996 (1989).
21. Kintner, C. R. & Brockes, J. P. *Nature* **308**, 67–69 (1984).
22. Mohun, T. J., Taylor, M. V., Garrett, N. & Gurdon, J. B. *EMBO J.* **8**, 1153–1161 (1989).
23. Krieg, P. A., Varnum, S. M., Wormington, W. M. & Melton, D. A. *Devl Biol.* **133**, 93–100 (1989).
24. Tapscott, S. J. *et al. Science* **242**, 405–411 (1988).
25. Krieg, P. A. & Melton, D. A. *Meth. Enzym.* **155**, 397–415 (1987).
26. Jackson, R. J. & Hunt, T. *Meth. Enzym.* **96**, 50–74 (1983).
27. Gurdon, J. B. *Meth. cell. Biol.* **16**, 125–139 (1977).

ACKNOWLEDGEMENTS. We thank D. Melton for pSP64-Xβm, J. Brockes for 12/101 antibody, and M. Taylor, C. Sharpe, J. Roche, J. Rashbass, B. McCabe, K. Kato and K. Kao for advice and comments on the manuscript. This work was supported by the Cancer Research Campaign and the Royal Society, of which N.D.H. is Challenor research fellow.

Enhancement of chilling tolerance of a cyanobacterium by genetic manipulation of fatty acid desaturation

Hajime Wada, Zoltan Gombos & Norio Murata*

National Institute for Basic Biology, Myodaiji, Okazaki 444, Japan

THE sensitivity (or tolerance) of plants to chilling determines their choice of natural habitat and also limits the worldwide production of crops. Although the molecular mechanism for chilling sensitivity has long been debated, no definitive conclusion has so far been reached about its nature. A probable hypothesis[1,2], however, is that chilling injury is initiated by phase transition of lipids of cellular membranes, as demonstrated for cyanobacteria, which serve as a model system for the plant cells[3,4]. Because the phase transition temperature depends on the degree of unsaturation of fatty acids of the membrane lipids[5], it is predicted that the chilling tolerance of plants can be altered by genetically manipulating fatty-acid desaturation by introducing double bonds into fatty acids of membrane lipids. Here we report the cloning of a gene for the plant-type desaturation (termed *desA*). The introduction of this gene from a chilling-resistant cyanobacterium, *Synechocystis* PCC6803, into a chilling-sensitive cyanobacterium, *Anacystis nidulans*, increases the tolerance of the recipient to low temperature.

A mutant in fatty-acid desaturation of membrane lipids of the transformable cyanobacterium, *Synechocystis* PCC6803, was isolated as described previously[6]. This mutant, termed Fad12, is defective in the activity of desaturation that introduces a second double bond at the Δ^{12} position of the C_{18} fatty acids of membrane lipids[6]. It grows much slower at low temperatures (such as 22 °C) than the wild type[6].

TABLE 1 Composition of major fatty acids of total membrane lipids in various strains of *Synechocystis* PCC6803

	Fatty acid (mol %)					
Strain	16:0	18:0	18:1 (9)	18:2 (6, 9)	18:2 (9, 12)	18:3 (6, 9, 12)
Wild type	58	1	8	t	12	17
Mutant (Fad12)	59	1	32	5	t	t
Transformant of Fad12 with *desA* (Bluescript/1.5 kbp)	59	2	9	1	12	14
Transformant of Fad12 with *desA* (pTZ19R/8 kbp)	59	1	8	t	12	15
Transformant of wild type with *desA*::Km^r (Bluescript/1.5 kbp::Km^r)	60	1	32	5	t	t

t, trace amount (<0.4%). Wild type, mutant (Fad12), and transformants of Fad12 with *desA* were grown photoautotrophically at 34 °C as described previously[13]. Transformant of wild type with *desA*::Km^r was cultivated in the same way but in the presence of 5 μg ml^{-1} kanamycin in the culture medium. The disrupted gene, *desA*::Km^r, was constructed by interrupting the *desA* in the cloned Bluescript/1.5 kbp at the *Hin*dIII site by the aminoglycoside 3′-phosphotransferase gene (the kanamycin-resistance (Km^r) cartridge) originating from the bacterial transposon Tn5 (ref. 16). Fatty acids of the total membrane lipids were analysed according to Sato and Murata[17]. The values are the means obtained in three independent experiments, and the deviation of values was within ±1.0%.

* To whom correspondence should be addressed.

a

```
   1 AATTCAGAAGCAATTCGGTTCCTGGTCAATGGCAACGTGTTATAAAAAGAAAAGTTTGTT
  61 TACCTGAGTATTAATTCCTAGGCACGGCAAACCTTGGCCGCTTTATAGCCCATGAATCCA
 121 TAAACAAAATCTGTCCGACCTTCCATTTGGAGATAAACCTTTATAA

 167 ATGACTGCCACGATTCCCCCGTTGACACCAACGGTAACGCCCAGCAATCCCGATCGCCCG
   1 MetThrAlaThrIleProProLeuThrProThrValThrProSerAsnProAspArgPro

 227 ATTGCGGATCTCAAACTACAAGACATCATTAAAACCCTGCCCAAGGAATGCTTCGAGAAA
  21 IleAlaAspLeuLysLeuGlnAspIleIleLysThrLeuProLysGluCysPheGluLys

 287 AAAGCGAGCAAAGCCTGGGCTTCTGTTTTGATTACCCTAGGGGCGATCGCCGTGGGCTAT
  41 LysAlaSerLysAlaTrpAlaSerValLeuIleThrLeuGlyAlaIleAlaValGlyTyr

 347 TTGGGCATTATTTATCTGCCCTGGTACTGCTTGCCCATTACCTGGATCTGGACAGGGACA
  61 LeuGlyIleIleTyrLeuProTrpTyrCysLeuProIleThrTrpIleTrpThrGlyThr

 407 GCCTTAACGGGGGCCTTCGTTGTCGGCCATGACTGTGGCCATCGCTCCTTTGCTAAAAAA
  81 AlaLeuThrGlyAlaPheValValGlyHisAspCysGlyHisArgSerPheAlaLysLys

 467 CGCTGGGTCAATGATTTAGTGGGACATATCGCTTTTGCTCCCCTCATCTACCCTTTCCAT
 101 ArgTrpValAsnAspLeuValGlyHisIleAlaPheAlaProLeuIleTyrProPheHis

 527 AGCTGGCGCCTACTCCACGACCACCATCACCTCCACACCAACAAAATTGAGGTTGATAAC
 121 SerTrpArgLeuLeuHisAspHisHisHisLeuHisThrAsnLysIleGluValAspAsn

 587 GCCTGGGATCCCTGGAGTGTGGAAGCTTTCCAAGCCAGCCCGGCGATCGTCCGGCTTTTT
 141 AlaTrpAspProTrpSerValGluAlaPheGlnAlaSerProAlaIleValArgLeuPhe

 647 TATCGGGCCATCCGGGGTCCCTTCTGGTGGACTGGTTCCATTTTCCATTGGAGCTTAATG
 161 TyrArgAlaIleArgGlyProPheTrpTrpThrGlySerIlePheHisTrpSerLeuMet

 707 CACTTCAAACTTTCCAACTTTGCCCAAAGGGACCGCAATAAAGTCAAATTATCCATTGCC
 181 HisPheLysLeuSerAsnPheAlaGlnArgAspArgAsnLysValLysLeuSerIleAla

 767 GTTGTCTTCCTGTTTGCGGCGATCGCCTTTCCTGCCCTAATTATCACCACAGGGGTGTGG
 201 ValValPheLeuPheAlaAlaIleAlaPheProAlaLeuIleIleThrThrGlyValTrp

 827 GGTTTCGTCAAATTTTGGCTAATGCCCTGGTTGGTGTATCACTTTTGGATGAGCACTTTT
 221 GlyPheValLysPheTrpLeuMetProTrpLeuValTyrHisPheTrpMetSerThrPhe

 887 ACCATTGTGCACCACACCATTCCCGAAATTCGTTTCCGTCCCGCCGCCGATTGGAGTGCC
 241 ThrIleValHisHisThrIleProGluIleArgPheArgProAlaAlaAspTrpSerAla

 947 GCCGAAGCCCAGTTAAATGGTACTGTTCACTGCGATTATCCCCGTTGGGTGGAAGTGCTC
 261 AlaGluAlaGlnLeuAsnGlyThrValHisCysAspTyrProArgTrpValGluValLeu

1007 TGCCATGACATCAACGTCCATATTCCCCACCACCTCTCCGTTGCCATCCCTTCCTATAAC
 281 CysHisAspIleAsnValHisIleProHisHisLeuSerValAlaIleProSerTyrAsn

1067 CTACGACTAGCCCACGGAAGTTTAAAAGAAAACTGGGGACCTTTTCTTTACGAGCGCACC
 301 LeuArgLeuAlaHisGlySerLeuLysGluAsnTrpGlyProPheLeuTyrGluArgThr

1127 TTTAACTGGCAATTAATGCAACAAATTAGTGGGCAATGTCATTTATATGACCCCGAACAT
 321 PheAsnTrpGlnLeuMetGlnGlnIleSerGlyGlnCysHisLeuTyrAspProGluHis

1187 GGCTACCGCACCTTCGGCTCCCTGAAAAAAGTTTAATACTGGGACAACTAGTAATTTTTG
 341 GlyTyrArgThrPheGlySerLeuLysLysVal***

1247 ACCCATGATTGGTCAGTAATTAACTTTGACTGATCCCCAGGGAGAGAAATACCAGATCAC
1307 AAATTAACTATCTTGGAATGCGGCCATCGAGCTGTATTCCTTTTTTCTTTTCTTGGTGAG
1367 GAAAAAAACTTTCTAAGTGGGCAGATGGGAGCGGTTCAGACTAGAAAGATCCACTCGGCC
1427 AAAATGAATGCTAAACGCAACCTGCATATTCTCCAAGGTTTTCACTAGCCAAGGTACCCC
1487 TTC
```

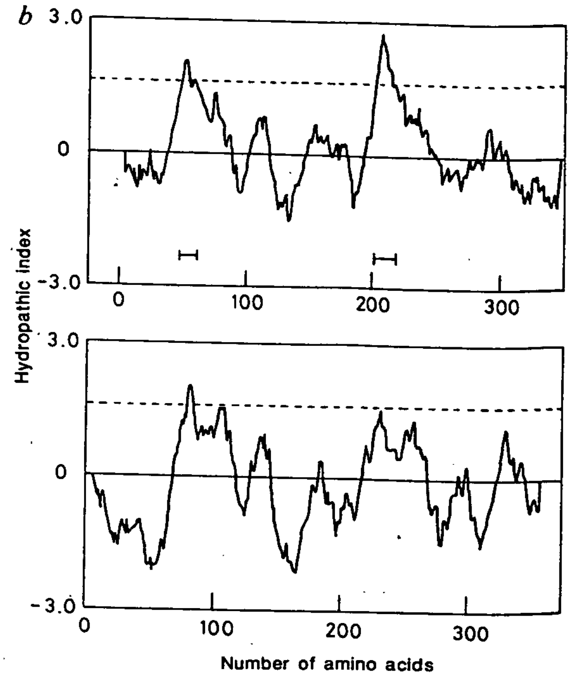


FIG. 1 *a*, Nucleotide sequence and deduced amino-acid sequence of *desA*, a gene for fatty-acid desaturation at the Δ^{12} position of fatty acids in *Synechocystis* PCC6803. The deduced amino-acid sequence is numbered with 1 for the first methionine. *b*, Hydropathy profiles of the deduced amino-acid sequences of the *desA* product and its putative membrane-spanning regions indicated by solid bars (top graph), and the stearoyl-CoA desaturase from rat liver[9] (bottom graph).

METHODS. Nucleotide sequence was determined by the dideoxy chain-termination method using double-stranded DNA templates[18]. The unidirectional deletion of the plasmid was performed according to the instructions of the manufacturer of the Bluescript DNA sequencing system (Stratagene Cloning Systems). Hydropathic index was calculated according to the algorithm of Kyte and Doolittle[19] for a window size of 19 amino-acid residues.

To clone a gene required for the desaturation at the Δ^{12} position, a genomic library of *Synechocystis* PCC6803 was constructed in plasmid pTZ19R. The genomic library was screened for clones capable of complementing Fad12 in the growth at low temperature and the desaturation at Δ^{12} position according to the *in situ* transformation method developed by Dzelzkalns and Bogorad[7]. A plasmid clone with an 8-kilobase pair (kbp) insert, termed pTZ19R/8 kbp, was isolated (Table 1). The homologous recombinations as described by Williams[8] between pTZ19R/8 kbp and the mutated gene of the chromosome of Fad12 may have taken place. The plasmid, pTZ19R/8 kbp, was digested with *Ava*I to obtain a 1.5-kbp fragment, which could also complement Fad12. The 1.5-kbp fragment was subcloned into a Bluescript plasmid (termed Bluescript/1.5 kbp), and its nucleotide sequence was determined. In only one of the six possible reading frames was there an open-reading frame. This was of 1,053 bp and corresponded to 351 amino-acid residues (Fig. 1*a*). The 1.5-kbp fragment also contained a 5′ upstream region of 166 bp and a 3′ downstream region of 270 bp. This gene (termed *desA*) encodes either a plant-type desaturase, which can introduce the second *cis* double bond at the Δ^{12} position of fatty acid bound to membrane glycerolipids, or a cofactor of this desaturase (see below). The hydropathy profile of the deduced amino-acid sequence of the *desA* product (Fig. 1*b*) is similar to that of the stearoyl-CoA desaturase from rat liver[9]. The *desA* gene product has two clusters of hydrophobic regions which are putative membrane-spanning domains (Fig. 1*b*). But, the sequence similarity between the *desA* product and stearoyl-CoA desaturase from rat liver is <30% at the nucleotide level and <10% at the amino-acid level.

We transformed the *Synechocystis* mutant, Fad12, with *desA* included in Bluescript/1.5 kbp and pTZ19R/8 kbp to examine whether the *desA* product is responsible for the fatty-acid desaturation. Table 1 shows that the wild type and the transformants contained high levels of 18:1 (9), 18:2 (9, 12) and 18:3 (6, 9, 12) fatty acids (fatty acids are represented by numbers of carbon atoms and double bonds, before and after a colon, respectively, and the positions of double bonds, counted from the carboxy terminus (Δ), are indicated by numbers in parantheses). In Fad12, 18:1 (9) and 18:2 (6, 9) significantly increased, whereas 18:2 (9, 12) and 18:3 (6, 9, 12) decreased to trace amounts. It is noteworthy that Fad12 lacked the fatty acids having the double bond at the Δ^{12} position, and that this double bond was recovered by transformation with *desA*. Similar changes in the desaturation of fatty acids at the Δ^{12} position were observed in all lipid classes, monogalactosyl diacylglycerol, digalactosyl diacylglycerol, sulphoquinovosyl diacylglycerol and phosphatidylglycerol (data not shown). To examine whether

the *desA* gene product is necessary for fatty-acid desaturation, we transformed the wild-type *Synechocystis* PCC6803 according to Williams[8] with a disrupted *desA*, which was produced by insertion of a kanamycin-resistance cartridge (Km^r). The transformant of wild type with *desA*::Km^r had the same fatty-acid composition as that of Fad12 (Table 1).

Another transformable cyanobacterium, *Anacystis nidulans* R2-SPc, was transformed with *desA* according to Kuhlemeier and van Arkel[10]. It is noteworthy that *A. nidulans* is a member of the group of cyanobacteria, which are completely defective in desaturation at the Δ^{12} position[11,12]. Bluescript/1.5 kbp was digested with *Sac*I to obtain a fragment containing the total sequence of the 1.5-kbp insert. This fragment was subcloned into pUC303 (ref. 10), a shuttle vector between *A. nidulans* and *Escherichia coli*, at the *Sac*I site of the streptomycin-resistance gene but in the opposite direction (termed pUC303/*desA*). The wild-type and the transformant with pUC303 alone (control) contained 16:0, 16:1 (9), 18:0 and 18:1 (9) as the principal fatty acids, indicating that this cyanobacterium can introduce only one double bond into the C_{16} and C_{18} fatty acids (Table 2). In the transformant with pUC303/*desA*, fatty acids having two double bonds, 16:2 (9, 12) and 18:2 (9, 12), emerged to significant levels at the expense of 16:1 (9) and 18:1 (9). The transformation of *A. nidulans* R2-SPc with the disrupted *desA* by the Km^r cartridge was also carried out as above. The transformant with pUC303/*desA*::Km^r had a fatty-acid composition similar to that of the wild type and the transformant with pUC303. The lipid class composition and the lipid-to-protein ratio were not affected by transformation with pUC303 and pUC303/*desA*. These observations demonstrate that the transformant with *desA* has acquired the desaturase activity in introducing the second double bond at the Δ^{12} position of fatty acids, and that the 166-bp upstream sequence contains the promoter region of this gene.

A. nidulans is sensitive to chilling temperature[3,4,13]. At growth temperature, both plasma and thylakoid membranes are in the

TABLE 2 Composition of major fatty acids of total membrane lipids in various strains of *Anacystis nidulans* R2-SPc

	Fatty acid (mol %)					
Strain	16:0	16:1 (9)	16:2 (9, 12)	18:0	18:1 (9)	18:2 (9, 12)
Wild type	51	36	0	3	6	0
Transformant with pUC303	51	37	0	3	5	0
Transformant with pUC303/*desA*	47	29	5	5	2	6
Transformant with pUC303/*desA*::Km^r	50	33	0	5	9	0

Wild type was grown photoautotrophically at 34 °C as described previously[13]. Transformants with pUC303 and pUC303/*desA* were grown in the same way as above but in the presence of 7.5 μg ml^{-1} chloramphenicol. Transformant with pUC303/*desA*::Km^r was grown as wild type but in the presence of 7.5 μg ml^{-1} chloramphenicol and 5 μg ml^{-1} kanamycin. The disrupted gene, pUC303/*desA*::Km^r, was constructed by interrupting *desA* in the cloned pUC303/*desA* at the *Bam*HI site by the Km^r cartridge. Fatty acids were analysed according to Sato and Murata[17]. The values are the means obtained in three independent experiments, and the deviation of values was within ±1.0%. Ten independently obtained transformants with *desA* gave the same result as above.

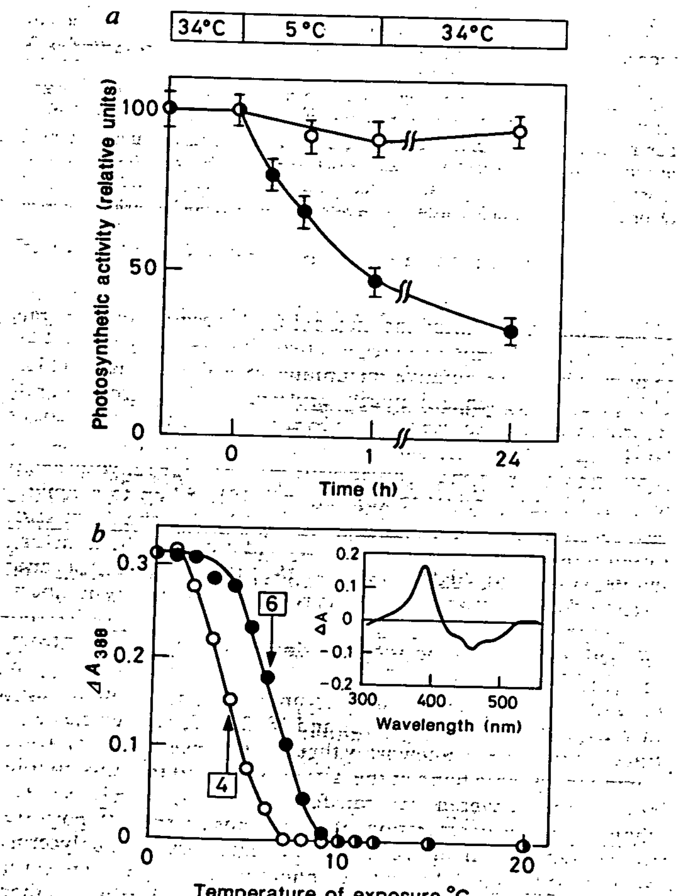


FIG. 2 Effects of low temperature on the photosynthetic activity and phase transition of membrane lipids in the transformants with and without *desA* of *Anacystis nidulans* R2-SPc. ●, Transformant with pUC303; ○, transformant with pUC303/*desA*. *a*, Effect of exposure to 5 °C on the photosynthetic activity. The activity before exposure to 5 °C corresponded to about 190 μmol O_2 per mg chlorophyll per h for both transformants. Each point represents the mean ± s.d. obtained in three independent experiments. *b*, Effect of exposure to low temperature for 20 min on the absorbance change at 388 nm. Inset: changes in absorption spectrum caused by exposure to 6 °C for 20 min in the transformant with pUC303 at a concentration corresponding to 15 μg chlorophyll ml^{-1}. Two independently obtained transformants gave essentially the same result.

METHODS. Cells grown at 34 °C were exposed to low temperature for 20 min in the dark. At an appropriate time the cells were rewarmed to 34 °C, and the photosynthetic activity (light-included oxygen evolution) and the absorption spectrum were measured according to Ono and Murata[13].

liquid-crystalline state. With decrease in growth temperature, the thylakoid membrane first goes into the phase-separated state only with reversible deterioration of photosynthesis. On further decrease in temperature, the plasma membrane enters the phase-separated state, in which leakage of the cytosolic solutes of low relative molecular mass into the medium irreversibly damages physiological activities[3,4].

When the wild type and transformant with pUC303 of *A. nidulans* R2-SPc grown at 34 °C were exposed to 5 °C for 60 min, more than 50% of photosynthetic activity was lost, and further decline of the activity continued after transferring them to 34 °C (Fig. 2*a*). By contrast, the transformant with pUC303/*desA* did not lose photosynthetic activity during the exposure to 5 °C for 60 min (Fig. 2*a*). On exposure to 2 °C for 60 min, the photosynthetic activity was decreased to 40% in the transformant with pUC303 and to 75% in the transformant with pUC303/*desA*. These observations demonstrate that chilling tolerance of *A. nidulans* R2-SPc was enhanced by transformation with *desA*.

The phase transition from the liquid-crystalline to the phase-separated state of the plasma membrane in intact cells of *A. nidulans* can be studied by changes in the absorption spectrum of carotenoids[3,13-15]. The phase transition of the plasma membranes of the transformants containing pUC303 or pUC303/*desA*, both grown at 34 °C, appeared in temperature ranges 8–4 °C with a midpoint at 6 °C, and 6–2 °C with a midpoint at 4 °C, respectively (Fig. 2*b*). The lowering in the phase transition temperature of the plasma membrane by *desA* can be regarded as resulting from the introduced desaturase activity, and this provides a molecular basis for the enhancement of chilling tolerance of this cyanobacterium.

The present study demonstrates that the chilling tolerance of cyanobacteria can be enhanced by genetic manipulation of fatty-acid desaturation. Because a similar mechanism could operate in the chilling injury of higher plants, it might be possible to improve their chilling tolerance by similarly manipulating fatty-acid desaturation. □

Received 24 April; accepted 2 July 1990.

1. Lyons, J. M. *A. Rev. Pl. Physiol.* **24**, 445–466 (1973).
2. Raison, J. K. *J. Bioenergetics* **4**, 285–309 (1973).
3. Murata, N. & Nishida, I. *The Biochemistry of Plants* Vol. 9 (Academic, Orlando, 1987).
4. Murata, N. *J. Bioenergetics Biomembranes* **21**, 61–75 (1989).
5. Chapman, D. *Q. Rev. Biophys.* **8**, 185–235 (1975).
6. Wada, H. & Murata, N. *Pl. Cell Physiol., Tokyo* **30**, 971–978 (1989).
7. Dzelzkalns, V. A. & Bogorad, L. *EMBO J.* **7**, 333–338 (1988).
9. Williams, J. G. K. *Meth. Enzym.* **167**, 766–778 (1988).
9. Thiede, M. A. *et al. J. biol. Chem.* **261**, 13230–13235 (1986).
10. Kuhlemeier, C. J. & van Arkel, G. A. *Meth. Enzym.* **153**, 199–215 (1987).
11. Kenyon, C. N. *J. Bacteriol.* **109**, 827–834 (1972).
12. Kenyon, C. N. *et al. Arch. Mikrobiol.* **83**, 216–236 (1972).
13. Ono, T. & Murata, N. *Pl. Physiol.* **67**, 176–181 (1981).
14. Omata, T. & Murata, N. *Pl. Cell Physiol., Tokyo* **24**, 1101–1112 (1983).
15. Gombos, Z. & Vigh, L. *Pl. Physiol.* **80**, 415–419 (1986).
16. Beck, E. *et al. Gene* **19**, 327–336 (1982).
17. Sato, N. & Murata, N. *Meth. Enzym.* **167**, 251–259 (1988).
18. Tabor, S. & Richardson, C. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4767–4771 (1987).
19. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105–132 (1982).

ACKNOWLEDGEMENTS. We thank T. Sakamoto for the transformation of *Synechocystis* PCC6803 with *desA*::Km^r, Dr N. Sato of the University of Tokyo for the gift of pUC303 and *A. nidulans* R2-SPc (obtained from Dr W. E. Borrias, University of Utrecht), and Professor T. Slabas (University of Durham) for reading the manuscript.

Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs

Gilbert Eriani, Marc Delarue*, Olivier Poch, Jean Gangloff & Dino Moras*

Laboratoires de Biochimie et de * Cristallographie Biologique, IBMC du CNRS, 15 rue Rene Descartes, 67084 Strasbourg, France

THE aminoacyl-transfer RNA synthetases (aaRS) catalyse the attachment of an amino acid to its cognate transfer RNA molecule in a highly specific two-step reaction. These proteins differ widely in size and oligomeric state, and have limited sequence homology. Out of the 18 known aaRS, only 9 (ref. 1), referred to as class I synthetases (GlnRS, TyrRS, MetRS, GluRS, ArgRS, ValRS, IleRS, LeuRS, TrpRS), display two short common consensus sequences ('HIGH' and 'KMSKS') which indicate, as observed in three crystal structures[2-4], the presence of a structural domain (the Rossman fold) that binds ATP. We report here the sequence of *Escherichia coli* ProRS, a dimer of relative molecular mass 127,402, which is homologous to both ThrRS and SerRS. These three latter aaRS share three new sequence motifs with AspRS, AsnRS, LysRS, HisRS and the β subunit of PheRS. These three motifs (motifs 1, 2 and 3), in a search through the entire data bank, proved to be specific for this set of aaRS (referred to as class II). Class II may also contain AlaRS and GlyRS, because these sequences have a typical motif 3. Surprisingly, this partition of aaRS in two classes is found to be strongly correlated on the functional level with the acylation occurring either on the 2′ OH (class I) or 3′ OH (class II) of the ribose of the last nucleotide of tRNA.

To clone the ProRS gene, a pool of partially digested *E. coli* DNA fragments was used to transform and complement the strain UQ27 (*proS27, argG, lac, thi*), a temperature-sensitive mutant defective in ProRS activity[5]. From complementing transformants we isolated several types of plasmids, whose insert sizes were reduced by limited *Sau*3A digestion, ligated into plasmid pUC18 vector, and selected in strain UQ27 at 43 °C. The resultant bacterial cells overproduced by 100 times the ProRS activity of wild-type cells. A 2.8-kilobase (kb) DNA fragment was subcloned into M13mp18, partially digested with exonuclease[6] and sequenced using modified T7 DNA polymerase[7]. The 2,795 base pairs (bp) that were sequenced contain an open reading frame encoding a protein of 572 residues whose relative molecular mass (M_r 63,701) is in good agreement with that estimated by SDS-PAGE (data not shown). The N-terminal protein sequence deduced from the primary structure was independently confirmed by sequencing the first 12 N-terminal residues of the purified ProRS. In addition, the *proS* messenger RNA 5′ termini were determined as described in Fig. 1. Analysis of the DNA region downstream of the TGA stop signal revealed a G+C-rich sequence of hyphenated dyad symmetry, centred on position +1,739 and followed by a run of seven T residues. This structural feature corresponds to a rho-independent termination signal, as indicated by the University of Winconsin Genetic Computer Group (UWGCG) TERMINATOR program using the algorithm of Brendel and Trifonov[8], which predicted a stop site for RNA polymerase at position +1,755 shown in Fig. 1.

The comparison of the ProRS sequence with other aaRS sequences showed extensive homologies with ThrRS (19.2% of strict identity and 44.8% of conservative substitutions as defined in the legend of Fig. 2, not considering the 110 amino-acid insert of ProRS) and with SerRS (13.7% of strict identity and 37.2% of conservative substitutions). An alignment of these three proteins is presented in Fig. 2. ProRS and ThrRS have both longer C-terminal domains than SerRS, whereas ThrRS has a long extension at its N terminus, as compared with ProRS and SerRS. Is this extension implicated in the autoregulation of the ThrRS translation, a unique feature of this aaRS? As already shown[9], this mechanism proceeds through an interaction between the ThrRS and a tRNA^{Thr} anticodon-like structure located upstream of the AUG initiation codon of *thrS*. More recently, however, several results[10] pointed to the existence, at the level of *thrS* operon, of a further domain interacting with ThrRS and whose

NOVEMBER 25, 1990 VOLUME 265 NUMBER 33

ISSN 0021-9258
JBCHA3 265(33) 20053–20713 (1990)

THE Journal of Biological Chemistry

Published by the American Society for Biochemistry and Molecular Biology

FOUNDED BY CHRISTIAN A HERTER
AND SUSTAINED IN PART BY THE CHRISTIAN A HERTER MEMORIAL FUND

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1990 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 265, No. 33, Issue of November 25, pp. 20144–20149, 1990
Printed in U.S.A.

The *OLE1* Gene of *Saccharomyces cerevisiae* Encodes the Δ9 Fatty Acid Desaturase and Can Be Functionally Replaced by the Rat Stearoyl-CoA Desaturase Gene*

(Received for publication, June 12, 1990)

Joseph E. Stukey‡, Virginia M. McDonough§, and Charles E. Martin¶

From the Nelson Biological Laboratory, Bureau of Biological Research, Rutgers University, Busch Campus, Piscataway, New Jersey 08855-1059

Strains of *Saccharomyces cerevisiae* bearing the *ole1* mutation are defective in unsaturated fatty acid (UFA) synthesis and require UFAs for growth. A previously isolated yeast genomic fragment complementing the *ole1* mutation has been sequenced and determined to encode the Δ9 fatty acid desaturase enzyme by comparison of primary amino acid sequence to the rat liver stearoyl-CoA desaturase. The *OLE1* structural gene encodes a protein of 510 amino acids (251 hydrophobic) having an approximate molecular mass of 57.4 kDa. A 257-amino acid internal region of the yeast open reading frame aligns with and shows 36% identity and 60% similarity to the rat liver stearoyl-CoA desaturase protein. This comparison disclosed three short regions of high consecutive amino acid identity (>70%) including one 11 of 12 perfect residue match. The predicted yeast enzyme contains at least four potential membrane-spanning regions and several shorter hydrophobic regions that align exactly with similar sequences in the rat liver protein. An *ole1* gene-disrupted yeast strain was transformed with a yeast-rat chimeric gene consisting of the promoter region and N-terminal 27 codons of *OLE1* fused to the rat desaturase coding sequence. Fusion gene transformants displayed near equivalent growth rates and modest lipid composition changes relative to wild type yeast control implying a significant conservation of Δ9 desaturase tertiary structure and efficient interaction b tw en the rat desaturase and yeast cytochrome b_5.

In animal and fungal cells, monounsaturated fatty acids are synthesized via an aerobic process from saturated fatty acid precursors by a microsomal membrane-bound three-component enzyme system involving cytochrome b_5, NADH-dependent cytochrome b_5 reductase, and the Δ9 fatty acid desaturase (1–3). This complex catalyzes the insertion of a double bond between carbons 9 and 10 of the saturated fatty acyl substrates, palmitoyl (16:0)- and stearoyl (18:0)-CoA, yielding the monoenoic products palmitoleic (16:1) or oleic (18:1) acids. Although higher eukaryotes contain polyunsaturated fatty acids in their membranes, either synthesized endogenously via Δ12 and Δ15 desaturase reactions or obtained from their diet, the Δ9 reaction accounts for all *de novo* unsaturated fatty acid (UFA)[1] production in *Saccharomyces cerevisiae* (4).

Isolation and characterization of fatty acid desaturase enzymes has proved difficult due to their extraordinary hydrophobic nature and tight association with membranes. Although fatty acid desaturation was first described using the yeast Δ9 desaturase system, only animal Δ9 enzymes have been successfully purified to homogeneity (5, 6). At a genetic level, only the DNA sequence for the rat liver and mouse adipocyte genes have been reported and analyzed (7, 8). Those genes were found to encode proteins with 92% identical amino acid sequences.

The Δ9 desaturase from rat liver has been most extensively characterized. It is a protein consisting of 358 amino acids of which 62% are hydrophobic (7). The functional enzyme has an obligate phospholipid requirement and contains one molecule of non-heme iron (5). Effects of chemical modification on enzyme function has suggested that arginyl and tyrosyl residues are involved in the binding of the negatively charged CoA moiety of the substrate and in the chelation of the iron prosthetic group, respectively (9). A truncated rat liver Δ9 enzyme missing 26 residues from the N terminus is also membrane-bound and functional (10).

Yeast mutants bearing the *ole1* allele require oleic acid for growth and were believed to produce a defective Δ9 desaturase suggesting that *OLE1* was the structural gene encoding the enzyme (11). Recently, we isolated and characterized a yeast genomic fragment containing the *OLE1* gene of *S. cerevisiae* (12). Replacement of the wild type gene in haploid cells with a disrupted form of that fragment resulted in a UFA-requiring, nonreverting phenotype.

In this paper we report the DNA sequence of the *S. cerevisiae OLE1* gene and compare the deduced amino acid sequence of the yeast Δ9 fatty acid desaturase with that of the rat liver stearoyl-CoA desaturase primary sequence. Although the proteins encoded are highly divergent, the rat Δ9 desaturase gene functions efficiently in *S. cerevisiae* in place of the native yeast gene. Furthermore, predicted structural features of the two proteins suggest a model for the topology of the Δ9 fatty acid desaturase in the ER membrane.

* This work was supported by National Science Foundation Grant DMB84-17802, Biomedical Research Group Grant RR-07058-21 from the Public Health Service, and by a grant from the Bureau of Biological Research Charles and Johanna Busch Memorial Fund. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J05676.

‡ Supported by a Charles and Johanna Busch predoctoral fellowship.

§ Supported by an Arthur McCallum predoctoral fellowship.

¶ To whom correspondence should be addressed. Tel.: 201-932-4081 or 201-873-2752.

[1] The abbreviations used are: UFA, unsaturated fatty acid; ORF, open reading frame; ER, endoplasmic reticulum; kb, kilobase(s).

MATERIALS AND METHODS

DNA Manipulations, Media, and Stain—All recombinant DNA manipulations were according to standard methods (13, 43). Plasmid amplifications and bacterial transformations were performed using either *Escherichia coli* strain HB101 or XL1 Blue (Stratagene). Yeast transformations were by the method of Ito *et al.* (14). Growth analysis was performed in synthetic dextrose medium supplemented with the appropriate amino acids (23). The genotype of yeast strain L8-14C is: α, *ole1*Δ::*LEU2, leu2-3, leu2-d112, ura3-52, his4*).

DNA Sequencing—Overlapping DNA fragments lying within the *OLE1* open reading frame were subcloned into pBluescript vectors (Stratagene) in two orientations for sequencing in either direction. Single-stranded DNA squencing templates were prepared by methods supplied by Stratagene. The M13(−20) primer was hybridized to single-stranded DNA templates and DNA sequencing was performed by the dideoxy chain termination method of Sanger *et al.* (15) using the modified T7 DNA polymerase, Sequenase (U. S. Biochemical Corp.). In two cases, *OLE1* internal oligonucleotides were synthesized to facilitate DNA sequence analysis.

DNA Sequence and Deduced Primary Sequence Analysis—*OLE1* DNA sequence and the deduced primary sequence analysis was performed using the Genetics Computer Group (GCG) sequence analysis software package (16). Amino acid sequence of the rat liver stearoyl-CoA desaturase was obtained from GenBank. Primary sequence comparison of the yeast and rat liver Δ9 desaturases was performed using the BestFit analysis program. Hydropathy analysis was according to Kyte and Doolittle (17).

Construction of Modified ole1 Alleles—Alleles *ole1–33* and *ole1–107*, containing stop codons in the 5′ region of the coding sequence, were constructed similarly. YEp352/OLE4.8 was partially digested with *Sal*I or *Nco*I, the cohesive ends were made blunt, and plasmids were religated. Following amplification in *E. coli*, plasmid samples were subject to restriction enzyme analysis. Candidates lacking the relevant restriction site were subject to DNA sequence analysis for verification.

Construction of Episomal and Centromeric Plasmids Bearing the Rat Liver Stearoyl-CoA Desaturase Gene—A 1.2-kb rat liver Δ9 desaturase cDNA fragment encoding residues 3–358, stop codon, and 136 base pairs of the 3′-untranslated region was removed from plasmid pDs3-358 (10) by digestion with *Bam*HI and *Sac*I and inserted into the multiple cloning site of episomal plasmid YEp352. A 1.0-kb yeast genomic fragment encompassing the promoter region, translation initiation codon, and the first 27 codons of the *OLE1* was isolated via *Hin*dIII/*Sal*I digestion and ligated in-frame with the rat desaturase fragment in YEp352. In this final construct, an eight-codon linker derived from the multiple cloning site regions of pUC8 and YEp352 separates the yeast N-terminal codons from the rat desaturase sequence. The predicted size of the fusion gene product is 391 amino acid residues. The yeast-rat fusion gene was then recovered via *Hin*dIII/*Dra*I digestion and ligated into YCp50 using *Hin*dIII and *Nru*I restriction sites. Plasmids bearing the fusion gene were amplified in *E. coli* strain XL1-Blue and used to transform the yeast *ole1* gene-disrupted strain L8-14C.

Lipid Isolation and Fatty Acid Analysis—Lipids were extracted from whole yeast cells by direct saponification (18). Fatty acid methyl esters were prepared by transmethylation with boron trifluoride (19) and analyzed by gas chromatography using a 30-meter capillary column SP-2330 (Supelco) in a Hewlett-Packard 5710A chromatograph as previously reported (12).

RESULTS AND DISCUSSION

General Features of the OLE1 Structural Gene—In a previous report it was shown that a cloned 4.8-kb *Hin*dIII yeast genomic fragment, but not two subclones of that fragment terminating at a central *Kpn*I region, complemented the *ole1* mutation of *S. cerevisiae* (12). From that *Kpn*I junction, overlapping subclones were used to "walk" through the observed open reading frame (ORF) in both directions yielding the sequence strategy presented in Fig. 1. Both strands were sequenced through the entire ORF without ambiguity.

The DNA and deduced amino acid sequence of *OLE1* and flanking nucleotide sequence is shown in Fig. 2. The ORF is 1530 nucleotides long. Translation of the entire ORF would produce a 510-amino acid polypeptide having an approximately molecular mass of 57.4 kDa containing 49.2% hydrophobic and 25.7% charged (10.0% acidic and 15.7% basic) amino acid residues. No consensus *N*-glycosylation sites are present in the deduced amino acid sequence of *OLE1* and the protein does not appear to contain a cleavable N-terminal signal sequence.

Yeast TATA promoter elements are commonly found 40–120 base pairs upstream from transcription initiation sites (20) with an average mRNA leader sequence of 52 nucleotides (21). The *OLE1* promoter region has two consensus TATA promoter elements (TATAAA and TATATA) located at positions −30 and −156 relative to the ORF. A transcription initiation event, directed from the TATATA element located at −156, could yield a transcript having features consistent with the above observations. However, transcription initiation directed from the TATAAA promoter element located at −30 could result in an atypically short, nontranslated leader sequence relative to the first in-frame ATG. Furthermore, there are three additional in-frame ATG codons within the first 400 base pairs of the *OLE1* ORF at positions 56, 61, and 116 that could also serve as potential translation start sites (see Fig. 1). Due to the close proximity of the first ATG codon to the TATA promoter element at −30 and comparison with the rat desaturase (discussed below) that showed no significant similarities in the first 140 amino acids, we were prompted to test for functional *OLE1* products initiating from these downstream sites. Two modified *ole1* alleles were constructed (see "Materials and Methods") that shifted the ORF and introduced translation stop codons at either position 33 (*ole1–33*) or 107 (*ole1–107*). Both in-frame stop codons were positioned before the next available ATG sequence. The *ole1* gene-disrupted yeast strain L8-14C, bearing the deletion allele *ole1*Δ::*LEU2*, was transformed with either of the above alleles on an episomal plasmid and tested for the ability to grow in the absence of exogenous UFAs. (Strains bearing this *ole1* allele were previously shown (12) to completely lack Δ9 desaturase activity as determined by product formation and have limited and finite growth potential (4–5 generations) in UFA-free medium.) In both cases the transformed strain grew only when UFAs were added to the growth medium, which is consistent with the first in-frame ATG codon functioning as the primary site of translation initiation.

Yeast and Rat Liver Δ9 Enzyme Amino Acid Analysis—A computer search of homologies to all current entries in GenBank/EMBL protein data bases identified a single data

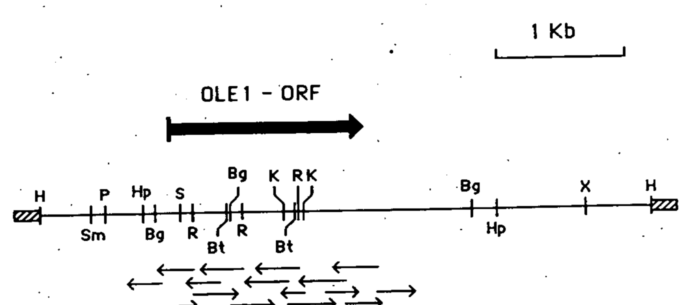


FIG. 1. ***OLE1* restriction map and sequencing strategy.** Primary restriction sites mapping the 4.8-kb yeast DNA fragment containing *OLE1*: *Bg*, *Bgl*II; *Bt*, *Bst*EII; *H*, *Hin*dIII; *Hp*, *Hpa*I; *K*, *Kpn*I; *R*, *Eco*RI; *S*, *Sal*I; *Sm*, *Sma*I; *P*, *Pst*I; and *X*, *Xho*I. The position and direction of the 1530-base-long ORF encoding the Δ9 enzyme is indicated by the *large arrow* above the map. *Small arrows* below the map indicate by size and direction the *OLE1* subclones used to sequence the entire ORF and flanking regions.

```
   1 ACACTCAACAAACCTTATCTAGTGCCCAACCAGGTGTGCTTCTACGAGTCTTGCTCACTC 60

  61 AGACACACCTATCCCTATTGTTACGGCTATGGGGATGGCACACAAAGGTGGAAATAATAG 120

 121 TAGTTAACAATATATGCAGCAAATCATCGGCTCCTGGCTCATCGAGTCTTGCAAATCAGC 180

 181 ATATACATATATATATGGGGGCAGATCTTGATTCATTTATTGTTCTATTTCCATCTTTCC 240

 241 TACTTCTGTTTCCGTTTATATTTTGTATTACGTAGAATAGAACATCATAGTAATAGATAG 300

 301 TTGTGGTGATCATATTATAAACAGCACTAAAACATTACAACAAAGATGCCAACTTCTGGA 360
                                                  MetProThrSerGly

 361 ACTACTATTGAATTGATTGACGACCAATTTCCAAAGGATGACTCTGCCAGCAGTGGCATT 420
     ThrThrIleGluLeuIleAspAspGlnPheProLysAspAspSerAlaSerSerGlyIle

 421 GTCGACGAAGTCGACTTAACGGAAGCTAATATTTTGGCTACTGGTTTGAATAAGAAAGCA 480
     ValAspGluValAspLeuThrGluAlaAsnIleLeuAlaThrGlyLeuAsnLysLysAla

 481 CCAAGAATTGTCAACGGTTTTGGTTCTTTAATGGGCTCCAAGGAAATGGTTTCCGTGGAA 540
     ProArgIleValAsnGlyPheGlySerLeuMetGlySerLysGluMetValSerValGlu

 541 TTCGACAAGAAGGGAAACGAAAAGAAGTCCAATTTGGATCGTCTGCTAGAAAAGGACAAC 600
     PheAspLysLysGlyAsnGluLysLysSerAsnLeuAspArgLeuLeuGluLysAspAsn

 601 CAAGAAAAAGAAGAAGCTAAAACTAAAATTCACATCTCCGAACAACCATGGACTTTGAAT 660
     GlnGluLysGluGluAlaLysThrLysIleHisIleSerGluGlnProTrpThrLeuAsn

 661 AACTGGCACCAACATTTGAACTGGTTGAACATGGTTCTTGTTTGTGGTATGCCAATGATT 720
     AsnTrpHisGlnHisLeuAsnTrpLeuAsnMetValLeuValCysGlyMetProMetIle

 721 GGTTGGTACTTCGCTCTCTCTGGTAAAGTACCTTTGCATTTAAACGTTTTCCTTTTCTCC 780
     GlyTrpTyrPheAlaLeuSerGlyLysValProLeuHisLeuAsnValPheLeuPheSer

 781 GTTTTCTACTACGCTGTCGGTGGTGTTTCTATTACTGCCGGTTACCATAGATTATGGTCT 840
     ValPheTyrTyrAlaValGlyGlyValSerIleThrAlaGlyTyrHisArgLeuTrpSer

 841 CACAGATCTTACTCCGCTCACTGGCCATTGAGATTATTCTACGCTATCTTCGGTTGTGCT 900
     HisArgSerTyrSerAlaHisTrpProLeuArgLeuPheTyrAlaIlePheGlyCysAla

 901 TCCGTTGAAGGGTCCGCTAAATGGTGGGGCCACTCTCACAGAATTCACCATCGTTACACT 960
     SerValGluGlySerAlaLysTrpTrpGlyHisSerHisArgIleHisHisArgTyrThr

 961 GATACCTTGAGAGATCCTTATGACGCTCGTAGAGGTCTATGGTACTCCCACATGGGATGG 1020
     AspThrLeuArgAspProTyrAspAlaArgArgGlyLeuTrpTyrSerHisMetGlyTrp

1021 ATGCTTTTGAAGCCAAATCCAAAATACAAGGCTAGAGCTGATATTACCGATATGACTGAT 1080
     MetLeuLeuLysProAsnProLysTyrLysAlaArgAlaAspIleThrAspMetThrAsp

1081 GATTGGACCATTAGATTCCAACACAGACACTACATCTTGTTGATGTTATTAACCGCTTTC 1140
     AspTrpThrIleArgPheGlnHisArgHisTyrIleLeuLeuMetLeuLeuThrAlaPhe

1141 GTCATTCCAACTCTTATCTGTGGTTACTTTTTCAACGACTATATGGGTGGTTTGATCTAT 1200
     ValIleProThrLeuIleCysGlyTyrPhePheAsnAspTyrMetGlyGlyLeuIleTyr

1201 GCCGGTTTTATTCGTGTCTTTGTCATTCAACAAGCTACCTTTTGCATTAACTCCATGGCT 1260
     AlaGlyPheIleArgValPheValIleGlnGlnAlaThrPheCysIleAsnSerMetAla

1261 CATTACATCGGTACCCAACCATTCGATGACAGAAGAACCCCTCGTGACAACTGGATTACT 1320
     HisTyrIleGlyThrGlnProPheAspAspArgArgThrProArgAspAsnTrpIleThr

1321 GCCATTGTTACTTTCGGTGAAGGTTACCATAACTTCCACCACGAATTCCCAACTGATTAC 1380
     AlaIleValThrPheGlyGluGlyTyrHisAsnPheHisHisGluPheProThrAspTyr

1381 AGAAACGCTATTAAGTGGTACCAATACGACCCAACTAAGGTTATCATCTATTTGACTTCT 1440
     ArgAsnAlaIleLysTrpTyrGlnTyrAspProThrLysValIleIleTyrLeuThrSer

1441 TTAGTTGGTCTAGCATACGACTTGAAGAAATTCTCTCAAAATGCTATTGAAGAAGCCTTG 1500
     LeuValGlyLeuAlaTyrAspLeuLysLysPheSerGlnAsnAlaIleGluGluAlaLeu

1501 ATTCAACAAGAACAAAAGAAGATCAATAAAAAGAAGGCTAAGATTAACTGGGGTCCAGTT 1560
     IleGlnGlnGluGlnLysLysIleAsnLysLysLysAlaLysIleAsnTrpGlyProVal

1561 TTGACTGATTTGCCAATGTGGGACAAACAAACCTTCTTGGCTAAGTCTAAGGAAAACAAG 1620
     LeuThrAspLeuProMetTrpAspLysGlnThrPheLeuAlaLysSerLysGluAsnLys

1621 GGTTTGGTTATCATTTCTGGTATTGTTCACGACGTATCTGGTTATATCTCTGAACATCCA 1680
     GlyLeuValIleIleSerGlyIleValHisAspValSerGlyTyrIleSerGluHisPro

1681 GGTGGTGAAACTTTAATTAAAACTGCATTAGGTAAGGACGCTACCAAGGCTTTCAGTGGT 1740
     GlyGlyGluThrLeuIleLysThrAlaLeuGlyLysAspAlaThrLysAlaPheSerGly

1741 GGTGTCTACCGTCACTCAAATGCCGCTCAAAATGTCTTGGCTGATATGAGAGTGGCTGTT 1800
     GlyValTyrArgHisSerAsnAlaAlaGlnAsnValLeuAlaAspMetArgValAlaVal

1801 ATCAAGGAAAGTAAGAACTCTGCTATTAGAATGGCTAGTAAGAGAGGTGAAATCTACGAA 1860
     IleLysGluSerLysAsnSerAlaIleArgMetAlaSerLysArgGlyGluIleTyrGlu

1861 ACTGGTAAGTTCTTTTAAGCATCACATTACAATAACAAAACTGCAACTACCATTAAAAAA 1920
     ThrGlyLysPhePheEnd

1921 AAATTGAAAAATCATAAATTAAAAAAAAAAAAATCAATTGAATTTTTTTTTTCATGATT 1980

1981 ACGTTTTGACATTTTTCCTTTTTTTTTCTCTTATTACGATTTACCTTTTTATTTATTTT 2040

2041 TTTCATTTTAGTATTTTATTCTTCGTTATTTATGTATAGAAATTTTCATTTTCATTTAGA 2100

2101 TTCAGATTTGGTTATCTTTTTTCATTATATATCTTTTGCACTAAGTTTCAGCTTAAGTTC 2160

2161 TATTTTTTATTTTTTTTTTCTGGGCCCTGGAGCAATAGATATGGGATGGCTTACTGCATC 2220

2221 TCTTCAAAATTTCACAGTCATGCTCACCCTTAAGTTCTCAACCTTTT 2267
```

FIG. 2. **Nucleotide and encoded amino acid sequence of the Δ9 fatty acid desaturase structural gene, *OLE1*.** Two consensus yeast TATA elements preceding the 1530-base-long ORF and the first four in-frame methionine-specific codons are *underlined*. An *OLE1* internal region of 258 amino acids displaying significant identity to the rat liver Δ9 enzyme is delimited by *asterisks*. Potential membrane-spanning regions are highlighted with *lines* above nucleotide and amino acid sequences.

base entry, the rat liver Δ9 desaturase, with significantly similarity to the *OLE1* gene product (Fig. 3). The aligned sequences show 36% identity and are greater than 60% similar over the region encompassing the C-terminal 260 amino acids of the shorter rat protein. No significant similarities exist over the N-terminal 141 amino acids of the yeast and 99 amino acids of the rat sequences. The yeast open reading frame extends 113 amino acids beyond the C-terminal end of the rat sequence. Within the region of high amino acid similarity there are three segments, having a minimum length of 10 residues, of very high identity (>70%) beginning at *OLE1* amino acid positions 156, 331, and 368. The most highly conserved of these is the first region where 17 of 23 identities are observed including one stretch containing 11 of 12 perfect matches.

The most conserved amino acid type within the compared region of the yeast protein is histidine with 10 of 14 (71.4%) residues in perfect alignment. Two other amino acid residues, proline and arginine, also show greater than 50% total identity. Arginine residues of the rat liver enzyme have been previously identified as being involved in substrate binding

```
yeast 141 VFLFSVFYYAVGGVSITAGYHRLWSHRSYSAHWPLRLFYAIFGCASVEGS 190
           .:|:::||| ::::|||| |||||||.|.|::|||:|. | .. ..:..
rat    99 TLLWGIFYYLISALGITAGAHRLWSHRTYKARLPLRIFLIIANTMAFQND 148

yeast 191 AKWWGHSHRIHHRYTDTLRDPYDARRGLWYSHMGWMLLKPNPKYKARA.. 238
           .  |::.|| ||::.:|  ||.:.|||:::||:||:|::.:|  |.::
rat   149 VYEWARDHRAHHKFSETHADPHNSRRGFFFSHVGWLLVRKHPAVKEKGGK 198

yeast 239 .DITDMTDDWTIRFQHRHYILLMLLTAFVIPTLICGYFFND.YMGGLIYA 286
            |:.|:..:. :.||:|.|   :|| .|::|||:. | :.: :: :|: .
rat   199 LDMSDLKAEKLVMFQRRYYKPGLLLMCFILPTLVPWYCWGETFLHSLFVS 248

yeast 287 GFIRVFVIQQATFCINSMAHYIGTQPFDDRRTPRDNWITAIVTFGEGYHN 336
          .|:|  :: .||: :|| ||..| .|:|..  .|:| :..:...|||:||
rat   249 TFLRYTLVLNATWLVNSAAHLYGYRPYDKNIQSRENILVSLGSVGEGFHN 298

yeast 337 FHHEFPTDYRNAIKWYQYDPTKVIIYLTSLVGLAYDLKKFSQNAIEEALI 386
          :||.|| ||...  :::.: |..:|    . :||||| ||.|..|:  | |
rat . 299 YHHAFPYDYSASEYRWHINFTTFFIDCMAALGLAYDRKKVSKAAV.LARI 347

yeast 387 QQEQKKINKKK 397
          .....  :|..
rat   348 KRTGDGSHKSS 358
```

FIG. 3. **Amino acid sequence comparison of the yeast and rat liver Δ9 fatty acid desaturases.** A 257-residue internal region of the yeast Δ9 enzyme is aligned with the rat liver stearoyl-CoA desaturase. Comparison was prepared by the GCG sequence analysis program BestFit. Identical residue matches are indicated by *connecting solid lines*. Two or one point between residues indicate decreasing amino acid similarity. Percent similarity value is based on the number of identical and two-point amino acid comparisons. Segments showing high identity (>70%) are indicated with *lines* above those regions.

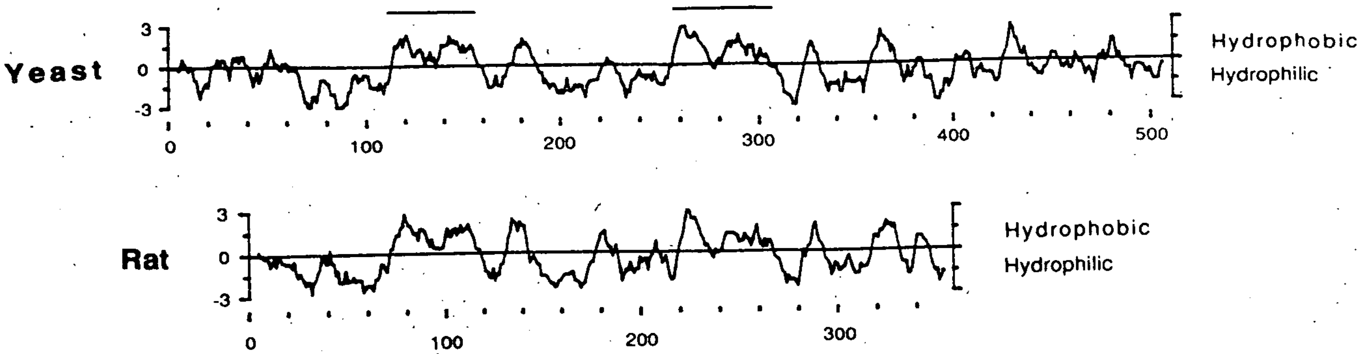


FIG. 4. **Hydropathy analysis of yeast and rat liver Δ9 enzymes.** Aligned Kyte-Doolittle (17) hydropathy profiles of the yeast and rat liver Δ9 desaturase proteins. The presumptive double membrane-spanning sequences are indicated by *bold lines* above those regions.

(9). Its significance as a highly conserved amino acid supports this finding. Although the role of histidyl residues in the fatty acid desaturases has not been examined, their highly conserved appearance also suggests an important contribution to enzyme function.

Structural Analysis and Proposed Topology of the Yeast Δ9 Enzyme—Striking similarities were also observed in the hydropathic characteristics of the two enzymes (Fig. 4). Both proteins contain two long hydrophobic regions (~50 residues) that could potentially form two membrane traversing loops, each consisting of two transmembrane segments. Chou-Fasman algorithms predict β-turn forming potential in the central portions of each loop in both the yeast and rat liver proteins. Inspection of the primary sequences at those sites reveals the presence of multiple helix-breaking amino acids that could serve to disrupt α-helical structure in order to form the looped structures. These hydrophobic regions are at identical positions in the aligned yeast and rat sequences. At least three smaller hydrophobic regions (each <7 amino acids) are also found at identical positions in the two proteins. The regions of high consecutive amino acid identity, however, are not within the hydrophobic sequences. The first region is located between the two "transmembrane loop regions," the second and third identity regions are located at the C-terminal part of the protein past the second "transmembrane loop." Neither extension of the N- and C-terminal domains of the yeast appears significantly hydrophobic and an examination of the amino acid distribution in those regions further suggests that they do not contribute to the integral membrane domains of the protein. A proposed model of the topology of the yeast protein in the ER membrane is given in Fig. 5. Assuming that the membrane-spanning regions are confined to the predicted hydrophobic sequences that are greater than 50 amino acids long, the arrangement places most of the protein on the cytosolic side of the ER membrane. Furthermore, all three regions of high consecutive identity would be located on that side of the membrane which is consistent with its proposed site of action (22).

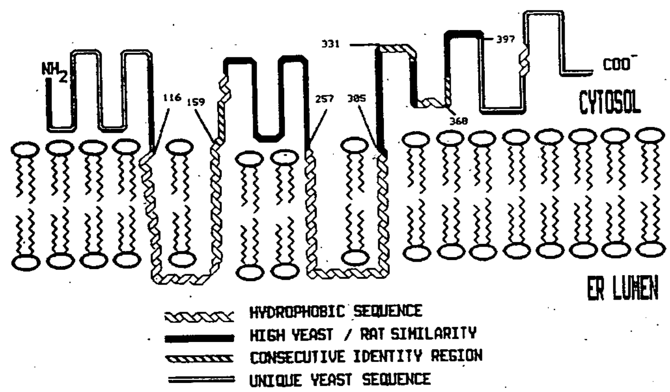


FIG. 5. **Model for the orientation of the yeast desaturase in the ER membrane.** The *numbers* identify amino acid positions in the yeast open reading frame.

Growth and Fatty Acid Content of Gene-disrupted Yeast Transformed with the Rat Liver Δ9 Desaturase—The significant sequence and predicted structural similarities observed between the yeast and rat Δ9 proteins prompted us to test whether the rat enzyme could functionally replace the yeast enzyme in *S. cerevisiae*, although there are additional residues at the N- and C-terminal ends of the yeast peptide sequence that are not found on the rat protein. A yeast-rat fusion gene was constructed (see "Materials and Methods") placing codons 3–358 of the rat gene in-frame with the initial 27 codons of the yeast gene and promoter sequences separated by an 8-codon linker region. This fusion gene was placed on a multicopy episomal and single copy centromere-based (CEN) vectors and introduced into the *ole1* gene-disrupted yeast strain, L8-14C. Fusion gene transformants were analyzed for growth and lipid composition relative to the same gene-disrupted strain transformed with the plasmid bearing native *OLE1* gene.

Yeast transformants bearing either the native *OLE1* or the yeast-rat fusion desaturase gene (two isolates) on an episomal plasmid were found cured of the UFA requirement and, surprisingly, showed identical growth rates (Fig. 6*A*) indicating significant conservation of Δ9 desaturase tertiary structure and an ability of the rat enzyme to interact with the yeast cytochrome b_5. In addition, because the rat protein is 113 residues shorter than the yeast desaturase at the C-terminal end and yet can functionally substitute for the yeast enzyme in *S. cerevisiae*, it appears that this extension of the yeast protein may be nonessential for catalytic functions. We cannot exclude the possibility, however, that the additional residues may be involved in other functions that influence its catalytic efficiency or optimize interactions with other components of the desaturase system.

An analysis of stationary phase cellular lipid compositions revealed, however, significant differences in the percentage of 16-carbon fatty acid species in the yeast-rat fusion gene transformants relative to the wild type control and, as a result, a modest decrease in the percent total UFA (Table I). The lower percentage of 16:1 and increased 16:0 species found in

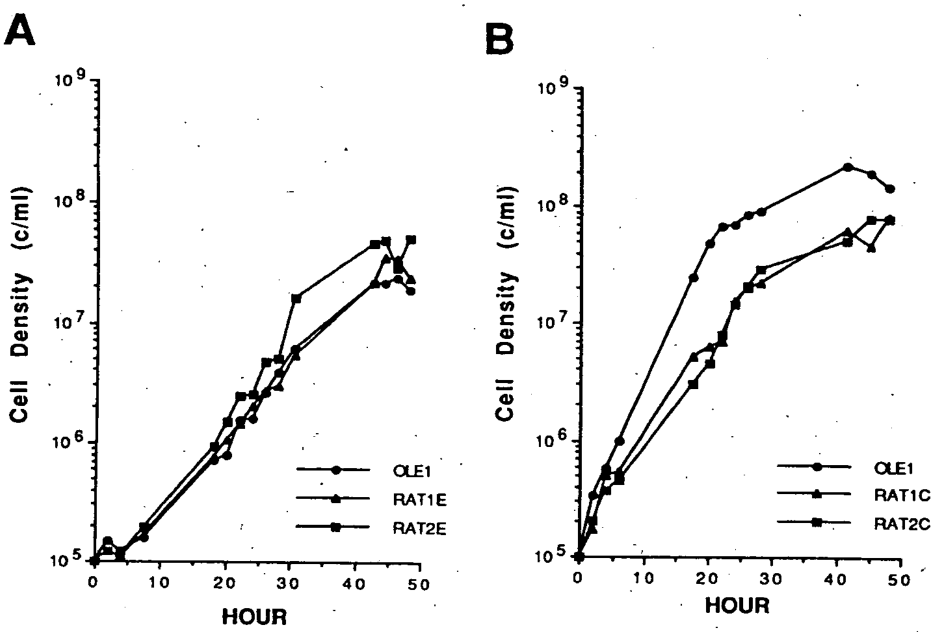


FIG. 6. **Growth characteristics of transformed *S. cerevisiae*.** Growth curves were determined for L8-14C transformants containing either the native *OLE1* gene or the yeast-rat Δ9 fusion gene on the episomal plasmid YEp352 (*A*) or the CEN plasmid YCp50 (*B*).

TABLE I

Fatty acid composition of transformed S. cerevisiae

Stationary phase L8-14C cells transformed with *OLE1* or the yeast-rat Δ9 chimeric gene on multiple (episomal) or single (CEN) copy number plasmids were harvested and cellular lipids analyzed as described under "Materials and Methods."

Plasmid type and transformant	Fatty acids					
	14:0	16:0	16:1	18:0	18:1	UFA
			%			
Episomal						
OLE1	1.95	21.92	41.70	4.65	29.77	71.47
RAT1E	1.40	33.00	28.77	4.61	32.21	60.98
RAT2E	1.41	30.94	27.42	4.73	35.51	62.93
CEN						
OLE1	1.16	18.38	34.05	9.10	37.32	71.36
RAT1C	5.58	40.67	21.22	8.70	23.82	45.04
RAT2C	7.23	39.71	15.83	10.89	26.35	42.18

those strains may reflect a preference of the rat Δ9 enzyme for the 18:0-CoA substrate over 16:0-CoA.

Although a yeast-rat Δ9 desaturase fusion gene is capable of functionally replacing the native *OLE1* of *S. cerevisiae* when present on a high copy number plasmid, a more stringent test of the efficiency of the rat protein in yeast would be to examine cells transformed with a single copy of the fusion gene. Cells containing the chimeric gene on CEN plasmid YCp50 showed growth rates that are reduced approximately 65% relative to wild type (Fig. 6*B*).

Similarly, the lipid composition of CEN plasmid-bearing yeast transformants differed markedly between those containing the chimeric gene and those containing the cloned yeast gene (Table I). The relative UFA levels were reduced approximately 38% in cells containing the rat gene coding sequence and the compensatory relative increase in saturated fatty acids resulted in a doubling of the 16:0 content and increased 14:0 levels, but no significant change in the level of 18:0. Thus, the yeast-rat Δ9 desaturase fusion gene can functionally replace the native *OLE1* of *S. cerevisiae*, although its action results in striking differences in cellular fatty acid compositions.

In previous studies using gene disruption and lipid analytical methods (12) we provided evidence suggesting that the *OLE1* gene encoded the yeast Δ9 fatty acid desaturase. The deduced *OLE1* amino acid sequence and physical comparisons of the yeast and rat liver proteins given here provide further proof that the *OLE1* locus contains the authentic structural gene for the desaturase. The aligned regions of consecutive identity between these two proteins from widely divergent sources suggests that they may represent conserved regions with similar function. The finding that the rat Δ9 fatty acid desaturase gene can complement *OLE1* in *S. cerevisiae* although the two proteins have only 36% identity suggests that there is conserved functional interaction among cytochrome b_5-mediated desaturase systems. Thus, ER-bound Δ9 enzymes from other organisms and possibly other cytochrome b_5-mediated desaturases, such as the Δ12 and Δ15, may also function in yeast.

Acknowledgment—We wish to thank Philipp Strittmatter for plasmids containing the rat stearoyl-CoA desaturase gene.

REFERENCES

1. Bloomfield, D. K., and Bloch, K. (1960) *J. Biol. Chem.* **235,** 337–345
2. Tamura, Y., Yoshida, Y., Sato, R., and Kumaoka, H. (1976) *Arch. Biochem. Biophys.* **175,** 284–294
3. Ohba, M., Sato, R., Yoshida, Y., Bieglmayer, C., and Ruis, H. (1979) *Biochim. Biophys. Acta* **572,** 352–362
4. Schweizer, E. (1984) in *Fatty Acid Metabolism and Its Regulation* (Numa, S., ed) pp. 59–83, Elsevier Scientific Publishing Co., New York
5. Strittmatter, P., Spatz, L., Corcoran, D., Rogers, M. J., Setlow, B., and Redline, R. (1974) *Proc. Natl. Acad. Sci. U. S. A.* **71,** 4565–4569
6. Prasad, M. R., and Joshi, V. C. (1979) *J. Biol. Chem.* **254,** 6362–6369
7. Thiede, M. A., Ozols, J., and Strittmatter, P. (1986) *J. Biol. Chem.* **261,** 13230–13235
8. Ntambi, J. M., Buhrow, S. A., Kaestner, K. H., Christy, R. J., Sibley, E., Kelly, T. J., Jr., and Lane, M. D. (1988) *J. Biol. Chem.* **263,** 17291–17300
9. Enoch, H. G., and Strittmatter, P. (1978) *Biochemistry* **17,** 4927–

4932
10. Strittmatter, P., Thiede, M. A., Hackett, C. S., and Ozols, J. (1988) *J. Biol. Chem.* **263,** 2532–2535
11. Resnick, M. A., and Mortimer, R. K. (1966) *J. Bacteriol.* **92,** 597–600
12. Stukey, J. E., McDonough, V. M., and Martin, C. E. (1989) *J. Biol. Chem.* **264,** 16537–16544
13. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
14. Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983) *J. Bacteriol.* **153,** 163–168
15. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
16. Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395
17. Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132
18. Kates, M. (1972) *Techniques of Lipidology*, Elsevier Scientific Publishing Co., New York
19. Morrison, W. R., and Smith, L. M. (1964) *J. Lipid Res.* **5,** 600–608
20. Struhl, K. (1989) *Annu. Rev. Biochem.* **58,** 1051–1077
21. Cigan, A. M., and Donahue, T. F. (1987) *Gene (Amst.)* **59,** 1–18
22. Jeffcoat, R., Brawn, P. R., Safford, R., and James, A. T. (1977) *Biochem. J.* **161,** 431–437
23. Sherman, F., Fink, G. R., and Hicks, J. B. (1982) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

ailable online!
:p://www.pnas.org

PROCEEDINGS

OF THE

NATIONAL ACADEMY OF SCIENCES

OF THE UNITED STATES OF AMERICA

April 1, 1997
Volume 94 / Number 7

INCLUDES: FROM THE ACADEMY FEATURING FRONTIERS OF SCIENCE

Plant Molecular Biology **27**: 293–300, 1993.

Isolation of a Δ^6-desaturase gene from the cyanobacterium *Synechocystis* sp. strain PCC 6803 by gain-of-function expression in *Anabaena* sp. strain PCC 7120

Avutu S. Reddy, Michael L. Nuccio, Lisa M. Gross and Terry L. Thomas *
Department of Biology, Texas A&M University, College Station, TX 77843, USA (author for correspondence)*

Received 9 December 1992; accepted in revised form 23 February 1993

Key words: fatty acid desaturation, C_{18} fatty acids, γ-linolenic acid, *Synechococcus* sp. strain PCC 7942

Abstract

The enzyme Δ^6-desaturase is responsible for the conversion of linoleic acid (18:2) to γ-linolenic acid (18:3γ). A cyanobacterial gene encoding Δ^6-desaturase was cloned by expression of a *Synechocystis* genomic cosmid library in *Anabaena*, a cyanobacterium lacking Δ^6-desaturase. Expression of the *Synechocystis* Δ^6-desaturase gene in *Anabaena* resulted in the accumulation of γ-linolenic acid (GLA) and octadecatetraenoic acid (18:4). The predicted 359 amino acid sequence of the *Synechocystis* Δ^6-desaturase shares limited, but significant, sequence similarity with two other reported desaturases. Analysis of three overlapping cosmids revealed a Δ^{12}-desaturase gene linked to the Δ^6-desaturase gene. Expression of *Synechocystis* Δ^6- and Δ^{12}-desaturases in *Synechococcus*, a cyanobacterium deficient in both desaturases, resulted in the production of linoleic acid and γ-linolenic acid.

Introduction

Appropriate control of lipid metabolism is critical to normal cellular and organismal function. In many instances, the number, position, and stereochemical orientation of carbon:carbon double bonds is critical to the biological activity of certain fatty acids. For example, there is considerable interest in the polyunsaturated C_{18} fatty acids: α-linolenic acid ($18{:}3^{\Delta 9,12,15}$) and γ-linolenic acid (GLA; $18{:}3^{\Delta 6,9,12}$). GLA is the result of desaturation of linoleic acid ($18{:}3^{\Delta 9,12}$) catalyzed by the enzyme Δ^6-desaturase. Consumption of vegetable oils rich in GLA may alleviate hypercholesterolemia and other clinical disorders which correlate with susceptibility to coronary heart disease [3]. The therapeutic benefits of dietary GLA may result from its being a precursor to arachidonic acid (20:4) and thus subsequently contributing to prostaglandin synthesis [26].

The nucleotide sequence data reported will appear in the EMBL GenBank and DDBJ Nucleotide Sequence Databases under the accession number L11421.

Most plant seed oils are deficient in GLA; therefore, we investigated the feasibility of obtaining a Δ^6-desaturase gene from a heterologous source which could, in turn, be used to transform plants to obtain seed oils containing GLA. The unicellular cyanobacterium *Synechocystis* PCC 6803 was chosen as a source for the Δ^6 desaturase because *Synechocystis* accumulates GLA to a level greater than 20% of the total fatty acid mass [16] and because the lipid composition of cyanobacteria is similar to that of higher plant chloroplasts [17]. Furthermore, unlike other prokaryotes, cyanobacteria have aerobic desaturases which make them good models for understanding lipid metabolism in higher plants [25].

With the exception of plant Δ^9-stearoyl acyl carrier protein desaturases, cyanobacterial, fungal, plant and animal desaturases are integral membrane proteins, a property that makes them difficult to purify and subsequently clone and characterize [1, 18, 21, 22, 24]. Therefore, we developed a molecular genetic strategy to isolate a Δ^6-desaturase gene from *Synechocystis* PCC 6803. A *Synechocystis* cosmid library was constructed and conjugated into wild-type *Anabaena* PCC 7120, a cyanobacterium deficient in Δ^6-desaturase, to identify gain-of-function *Anabaena* transconjugants that produce GLA and therefore contain a functional *Synechocystis* Δ^6-desaturase gene. With this approach, we cloned a Δ^6-desaturase gene from *Synechocystis* and verified its expression in another cyanobacterium, *Synechococcus* PCC 7942.

Materials and methods

Strains and culture conditions

Synechocystis PCC 6803 was obtained from the American Type Culture Collection. *Anabaena* PCC 7120 and *Synechococcus* PCC 7942 were kindly provided by Dr James Golden and Dr Susan Golden, respectively (Department of Biology, Texas A&M University). These strains were grown photoautotropically at 30 °C in BG-11 medium [19] under illumination of incandescent lamps (60 μE m^{-2} s^{-1}). Cosmids and plasmids were selected and propagated in *Escherichia coli* strain DH5α on LB medium supplemented with antibiotics at standard concentrations [15].

Construction of Synechocystis *cosmid genomic library*

Total genomic DNA from *Synechocystis* PCC 6803 was partially digested with *Sau*3AI and fractionated on a sucrose gradient [2]. Fractions containing 30 to 40 kb DNA fragments were selected and ligated into the dephosphorylated *Bam* HI site of the cosmid vector, pDUCA7 [4]. The ligated DNA was packaged *in vitro* [2], and packaged phage were propagated in *E. coli* DH5αmcr$^-$ containing the helper plasmid, pRL528 encoding *Ava* I and *E. coli* 47II methylases [10]. A total of 1152 colonies were randomly isolated and individually maintained in twelve 96 well microtiter plates.

Conjugation of Synechocystis *cosmid library into* Anabaena

Anabaena cells were grown to mid-log phase in BG-11 liquid medium, washed and resuspended in the same medium to a final concentration of ca. 2×10^8 cells per ml. A mid-log phase culture of *E. coli* containing the RP4 plasmid [5, 10] grown in LB containing 50 μg ampicillin per ml was washed and resuspended in fresh LB medium. *Anabaena* cells were then mixed with *E. coli* containing RP4 and spread evenly on BG-11 plates containing 5% LB. The cosmid genomic library was replica plated onto LB plates containing 50 μg kanamycin and 17.5 μg chloramphenicol per ml and was subsequently patched onto BG-11 plates containing *Anabaena* and *E. coli* carrying the RP4 plasmid. After 24 h of incubation at 30 °C, neomycin was underlaid to a final concentration of 30 μg/ml and incubation at 30 °C was continued until transconjugants appeared [10].

Fatty acid analysis

Wild-type and transgenic cyanobacterial cultures were grown as described [19], harvested by centrifugation, and washed twice with distilled water. Fatty acid methyl esters were prepared from these cultures [8] and were analyzed by gas-liquid chromatography (GLC) using a Tracor-560 equipped with a hydrogen flame ionization detector and capillary column (30 m × 0.25 mm bonded FSOT Superox II; Alltech Associates, IL). Retention times and co-chromatography of standards (obtained from Sigma Chemical Co.) were used for identification of fatty acids. The average fatty acid composition was determined as the ratio of peak area of each C_{18} fatty acid normalized to $C_{17:0}$ internal standard.

DNA sequence analysis

Standard molecular biology techniques were performed as described [2, 15]. Dideoxy sequencing [20] of pBS1.8 was performed with Sequenase (United States Biochemical) on both strands using specific oligonucleotide primers synthesized by the Advanced DNA Technologies Laboratory (Biology Department, Texas A&M University). DNA sequence analysis was done with the GCG (Madison, WI) software [9].

Results

Gain-of-function expression of GLA in Anabaena

Anabaena PCC 7120, a filamentous cyanobacterium, is deficient in GLA but contains significant amounts of linoleic acid, the precursor for GLA (Fig. 2; Table 1). A *Synechocystis* cosmid library was conjugated into *Anabaena* PCC 7120 to identify transconjugants that produce GLA. Individual transconjugants were isolated after conjugation and grown in 2 ml BG11N$^+$ liquid medium with 15 μg neomycin per ml. Fatty acid methyl esters were prepared from cultures containing pools of ten transconjugants and analyzed by GLC; representative GLC profiles are shown in Fig. 2. Two pools (of 25 pools representing 250 transconjugants) were identified that produced GLA. Individual transconjugants of each GLA positive pool were analyzed for GLA production; two independent transconjugants, *AS13* and *AS75*, one from each pool, were identified that expressed significant levels of GLA and which contained cosmids, cSy13 and cSy75, respectively (Fig. 1). These cosmids overlap in a region approximately 7.5 kb in length. A 3.5 kb *Nhe* I fragment of cSy75 was recloned in the vector pDUCA7 to create pSy75-3.5 and transferred to *Anabaena* resulting in gain-of-function expression of GLA (Table 1).

Two *Nhe* I/*Hin*d III subfragments (1.8 and 1.7 kb) of the 3.5 kb *Nhe* I fragment of pSy75-3.5 were subcloned into pBluescript (Fig. 1) for sequencing. Subsequently, both subfragments were transferred into a conjugal expression vector, pAM542 (T.S. Ramasubramanian and J. Golden, personal communication), in both forward and reverse orientations with respect to a cyanobacterial *rbcLS* promoter and were introduced into *Anabaena* by conjugation. Transconjugants containing the 1.8 kb fragment in the forward orientation (pAM542-1.8F) produced significant quantities of GLA and octadecatetraenoic acid (Fig. 2;

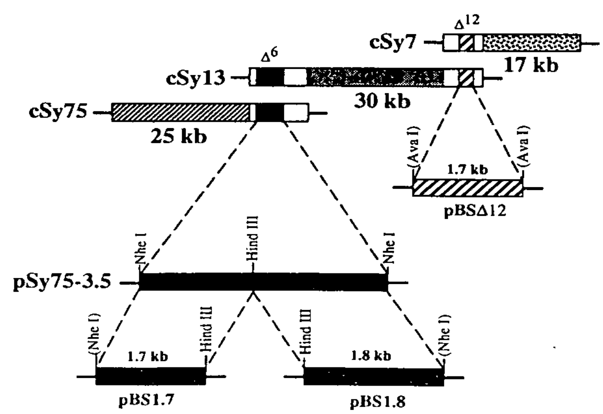


Fig. 1. Maps of cosmid cSy75, cSy13 and cSy7 with overlapping regions and subclones. The origin of subclones of cSy75, pSy75-3.5 and cSy7 are indicated by the dashed diagonal lines. Restriction sites that have been inactivated are in parenthesis.

Table 1. Composition of C_{18} fatty acids in wild-type and transgenic cyanobacteria.

Strain	Fatty acid (%)					
	18:0	18:1	18:2	18:3(α)	18:3(γ)	18:4
Wild type						
Synechocystis sp. PCC 6803	13.6	4.5	54.5	0	27.3	0
Anabaena sp. PCC 7120	2.9	24.8	37.1	35.2	0	0
Synechococcus sp. PCC 7942	20.6	79.4	0	0	0	0
Anabaena transconjugants						
cSy75	3.8	24.4	22.3	9.1	27.9	12.5
pSy75-3.5	4.3	27.6	18.1	3.2	40.4	6.4
pAM542-1.8F	4.2	13.9	12.1	19.1	25.4	25.4
pAM542-1.8R	7.7	23.1	38.4	30.8	0	0
pAM542-1.7F	2.8	27.8	36.1	33.3	0	0
pAM542-1.7R	2.8	25.4	42.3	29.6	0	0
Synechococcus transformants						
pAM854	27.8	72.2	0	0	0	0
pAM854-Δ12	4.0	43.2	46.0	0	0	0
pAM854-Δ6	18.2	81.8	0	0	0	0
pAM854-Δ6 & Δ12	42.7	25.3	19.5	0	16.5	0

18:0, stearic acid; 18:1, oleic acid; 18:2, linoleic acid; 18:3(α), α-linolenic acid; 18:3(γ), γ-linolenic acid; 18:4, octadecatetraenoic acid.

Table 1). Transconjugants containing other constructs, either reverse oriented 1.8 kb fragment or forward and reverse oriented 1.7 kb fragment, did not produce detectable levels of GLA (Table 1).

Figure 2 compares the C_{18} fatty acid profile of an extract from wild type *Anabaena* (Fig. 2A) with that of transgenic *Anabaena* containing the 1.8 kb fragment of pSy75-3.5 in the forward orientation (Fig. 2B). GLC analysis of fatty acid methyl esters from pAM542-1.8F revealed a peak with a retention time identical to that of an authentic GLA standard. Analysis of this peak by gas chromatography-mass spectrometry (GC-MS) confirmed that it had the same mass fragmentation pattern as a GLA reference sample (data not shown).

Two genes involved in C_{18} fatty acid biosynthesis are linked

We isolated a third cosmid, cSy7 containing a Δ^{12}-desaturase gene by screening the *Synechocystis* genomic library with an oligonucleotide synthesized from the published *Synechocystis* Δ^{12}-desaturase gene sequence [25]. We identified a 1.7 kb *Ava* I fragment from this cosmid containing the Δ^{12}-desaturase gene and subcloned it into pBluescript to create pBSΔ12 (Fig. 1). We then used this probe to demonstrate that cSy13 not only contains a Δ^{6}-desaturase gene but also a Δ^{12}-desaturase gene (Fig. 1). Genomic filter hybridizations further showed that both the Δ^{6}- and Δ^{12}-desaturase genes are unique in the *Synechocystis* genome indicating that two functional genes involved in C_{18} fatty acid desaturation are linked in the *Synechocystis* genome.

Sequence analysis and comparison with other desaturases

The nucleotide sequence of the 1.8 kb fragment of pSy75-3.5 including the functional Δ^{6}-desaturase gene was determined. An open reading frame encoding a polypeptide of 359 amino acids was

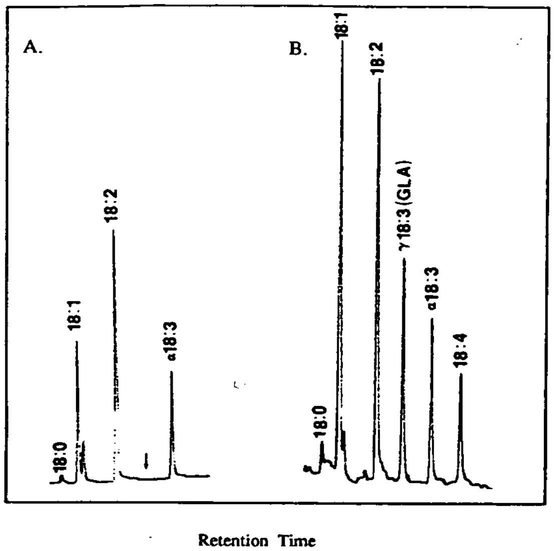


Fig. 2. GLC analysis of fatty acids of wild-type and transgenic *Anabaena*. C_{18} fatty acid methyl esters are shown. A. *Anabaena* wild type (arrow indicates migration time of GLA). B. Transconjugant of *Anabaena* with pAM542-1.8F. GLA, γ-linolenic acid; 18:4, octadecatetraenoic acid. Peaks were identified by comparing the elution times with known standards of fatty acid methyl esters and were confirmed by GC-MS.

identified (Fig. 3). It shares limited, but significant, amino acid sequence similarity with Δ^{15}-desaturase from *Brassica napus* [1] and Δ^{12}-desaturase [25]. A Kyte-Doolittle hydropathy analysis [14] identified two regions of hydrophobic amino acids that could represent transmembrane domains (Fig. 4A); furthermore, the hydropathic profile of the Δ^{6}-desaturase is similar to that of the *Synechocystis* Δ^{12}-desaturase gene (Fig. 4B; [25]), Δ^{9}-desaturase (not shown [23]) and Δ^{15}-desaturase (not shown [1]).

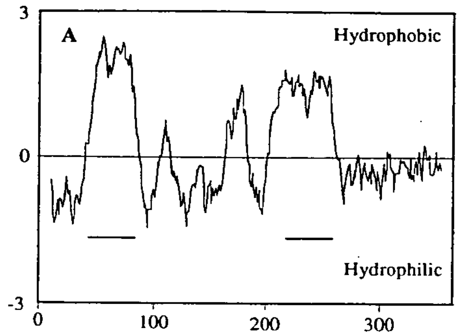
```
      AAGCTTCACTTCGGTTTTATATTGTGACCATGGTTCCCAGGCATCTGCTCTAGGGAGT -241
     TTTTCCGCTGCCTTTAGAGAGTATTTTCTCCAAGTCGGCTAACTCCCCCATTTTTAGGCA -181
     AAATCATATACAGACTATCCCAATATTGCCAGAGCTTTGATGACTCACTGTAGAAGGCAG -121
     ACTAAAATTCTAGCAATGGACTCCCAGTTGGAATAAATTTTTAGTCTCCCCCGGCGCTGG -61
     AGTTTTTTTGTAGTTAATGGCGGTATAATGTGAAAGTTTTTTATCTATTTAAATTTATAA -1
     ATG CTA ACA GCG GAA AGA ATT AAA TTT ACC CAG AAA CGG GGG TTT  45
   1 M   L   T   A   E   R   I   K   F   T   Q   K   R   G   F
     CGT CGG GTA CTA AAC CAA CGG GTG GAT GCC TAC TTT GCC GAG CAT  90
  16 R   R   V   L   N   Q   R   V   D   A   Y   F   A   E   H
     GGC CTG ACC CAA AGG GAT AAT CCC TCC ATG TAT CTG AAA ACC CTG  135
  31 G   L   T   Q   R   D   N   P   S   M   Y   L   K   T   L
     ATT ATT GTG CTC TGG TTG TTT TCC GCT TGG GCC TTT GTG CTT TTT  180
  46 I   I   V   L   W   L   F   S   A   W   A   F   V   L   F
     GCT CCA GTT ATT TTT CCG GTG CGC CTA CTG GGT TGT ATG GTT TTG  225
  61 A   P   V   I   F   P   V   R   L   L   G   C   M   V   L
     GCG ATC GCC TTG GCG GCC TTT TCC TTC AAT GTC GGC CAC GAT GCC  270
  76 A   I   A   L   A   A   F   S   F   N   V   G   H   D   A
     AAC CAC AAT GCC TAT TCC TCC AAT CCC CAC ATC AAC CGG GTT CTG  315
  91 N   H   N   A   Y   S   S   N   P   H   I   N   R   V   L
     GGC ATG ACC TAC GAT TTT GTC GGG TTA TCT AGT TTT CTT TGG CGC  360
 106 G   M   T   Y   D   F   V   G   L   S   S   F   L   W   R
     TAT CGC CAC AAC TAT TTG CAC CAC ACC TAC ACC AAT ATT CTT GGC  405
 121 Y   R   H   N   Y   L   H   H   T   Y   T   N   I   L   G
     CAT GAC GTG GAA ATC CAT GGA GAT GGC GCA GTA CGT ATG AGT CCT  450
 136 H   D   V   E   I   H   G   D   G   A   V   R   M   S   P
     GAA CAA GAA CAT GTT GGT ATT TAT CGT TTC CAG CAA TTT TAT ATT  495
 151 E   Q   E   H   V   G   I   Y   R   F   Q   Q   F   Y   I
     TGG GGT TTA TAT CTT TTC ATT CCC TTT TAT TGG TTT CTC TAC GAT  540
 166 W   G   L   Y   L   F   I   P   F   Y   W   F   L   Y   D
     GTC TAC CTA GTG CTT AAT AAA GGC AAA TAT CAC GAC CAT AAA ATT  585
 181 V   Y   L   V   L   N   K   G   K   Y   H   D   H   K   I
     CCT CCT TTC CAG CCC CTA GAA TTA GCT AGT TTG CTA GGG ATT AAG  630
 196 P   P   F   Q   P   L   E   L   A   S   L   L   G   I   K
     CTA TTA TGG CTC GGC TAC GTT TTC GGC TTA CCT CTG GCT CTG GGC  675
 211 L   L   W   L   G   Y   V   F   G   L   P   L   A   L   G
     TTT TCC ATT CCT GAA GTA TTA ATT GGT GCT TCG GTA ACC TAT ATG  720
 226 F   S   I   P   E   V   L   I   G   A   S   V   T   Y   M
     ACC TAT GGC ATC GTG GTT TGC ACC ATC TTT ATG CTG GCC CAT GTG  765
 241 T   Y   G   I   V   V   C   T   I   F   M   L   A   H   V
     TTG GAA TCA ACT GAA TTT CTC ACC CCC GAT GGT GAA TCC GGT GCC  810
 256 L   E   S   T   E   F   L   T   P   D   G   E   S   G   A
     ATT GAT GAC GAG TGG GCT ATT TGC CAA ATT CGT ACC ACG GCC AAT  855
 271 I   D   D   E   W   A   I   C   Q   I   R   T   T   A   N
     TTT GCC ACC AAT AAT CCC TTT TGG AAC TGG TTT TGT GGC GGT TTA  900
 286 F   A   T   N   N   P   F   W   N   W   F   C   G   G   L
     AAT CAC CAA GTT ACC CAC CAT CTT TTC CCC AAT ATT TGT CAT ATT  945
 301 N   H   Q   V   T   H   H   L   F   P   N   I   C   H   I
     CAC TAT CCC CAA TTG GAA AAT ATT ATT AAG GAT GTT TGC CAA GAG  990
 316 H   Y   P   Q   L   E   N   I   I   K   D   V   C   Q   E
     TTT GGT GTG GAA TAT AAA GTT TAT CCC ACC TTC AAA GCG GCG ATC  1035
 331 F   G   V   E   Y   K   V   Y   P   T   F   K   A   A   I
     GCC TCT AAC TAT CGC TGG CTA GAG GCC ATG GGC AAA GCA TCG TGA  1080
 346 A   S   N   Y   R   W   L   E   A   M   G   K   A   S   *
     CATTGCCTTGGGATTGAAGCAAAATGGCAAAATCCCTCGTAAATCTATGATCGAAGCCT  1139
     TTCTGTTGCCCGCCGACCAAATCCCCGATGCTGACCAAAGGTTGATGTTGGCATTGCTC  1248
     CAAACCCACTTTGAGGGGGTTCATTGGCCGCAGTTTCAAGCTGACCTAGGAGGCAAAGA  1307
     TTGGGTGATTTTGCTCAAATCCGCTGGGATATTGAAAGGCTTCACCACCTTTGGTTTCT  1366
     ACCCTGCTCAATGGGAAGGACAAACCGTCAGAATTGTTTATTCTGGTGACACCATCACC  1425
     GACCCATCCATGTGGTCTAACCCAGCCCTGGCCAAGGCTTGGACCAAGGCCATGCAAAT  1484
     TCTCCACGAGGCTAGGCCAGAAAAATTATATTGGCTCCTGATTTCTTCCGGCTATCGCA  1543
     CCTACCGATTTTTGAGCATTTTTGCCAAGGAATTCTATCCCCACTATCTCCATCCCACT  1602
     CCCCCGCCTGTACAAAATTTTATCCATCAGCTAGC                          1637
```

Fig. 3. Nucleotide and predicted amino acid sequences of the *Synechocystis* Δ^{6}-desaturase. Amino acid residues are numbered on the left; nucleotide positions are numbered on the right.

Transformation of Synechococcus *with Δ^{6}- and Δ^{12}-desaturase genes*

The unicellular cyanobacterium *Synechococcus* PCC 7942 is deficient in both linoleic acid and GLA [16]. We cloned Δ^{12} and Δ^{6}-desaturase genes individually and together into pAM854 [6], a shuttle vector that contains sequences necessary for the integration of foreign DNA into the genome of *Synechococcus* [11]. *Synechococcus* was transformed with these gene constructs and colonies were selected [6]. Fatty acid methyl esters were prepared from transgenic *Synechococcus* and analyzed by GLC.

Table 1 shows that the principal C_{18} fatty acids of wild-type *Synechococcus* are stearic acid (18:0)



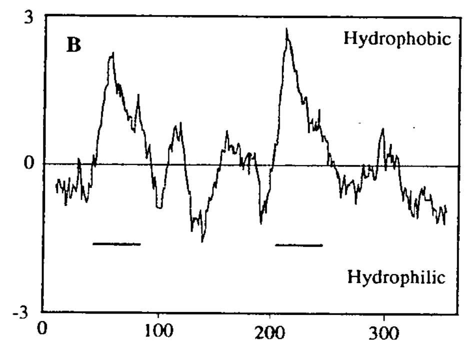


Fig. 4. Hydropathy profiles of (A) Δ^6- and (B) Δ^{12}-desaturase from *Synechocystis*. The Kyte and Doolittle algorithm in the GCG sequence analysis software was used to predict relative hydrophobicity in the predicted polypeptides of Δ^6-desaturase and Δ^{12}-desaturase [9]. Putative membrane-spanning regions are indicated by solid bars.

and oleic acid (18:1). *Synechococcus* transformed with pAM854-Δ12 expressed linoleic acid (18:2) in addition to the principal fatty acids. Transformants with pAM854-Δ6&Δ12 produced both linoleate and GLA (Table 1). These results indicate that *Synechococcus* containing both Δ^{12}- and Δ^6-desaturase genes has gained the capability of introducing a second double bond at the Δ^{12} position and a third double bond at the Δ^6 position of C_{18} fatty acids. However, no changes in fatty acid composition were observed in the transformant containing pAM854-Δ6 indicating that in the absence of substrate synthesized by the Δ^{12}-desaturase, the Δ^6-desaturase is inactive; whether Δ^6-desaturase requires two double bonds or only a double bond at the Δ^{12} position is not clear. This experiment further confirms that the 1.8 kb *Nhe* I/*Hind* III fragment (Fig. 1) contains both coding and promoter regions of the *Synechocystis* Δ^6-desaturase gene.

Discussion

We used a gain-of-function approach to identify a cyanobacterial gene encoding an enzyme involved in fatty acid metabolism. The enzyme Δ^6-desaturase is required for the conversion of linoleic acid ($18{:}2^{\Delta 9,12}$) to γ-linolenic acid ($18{:}3^{\Delta 6,9,12}$) or GLA. Conjugation of a *Synechocystis* PCC 6803 cosmid library into the filamentous cyanobacterium *Anabaena*, which lacks GLA but does contain linoleic acid, the precursor to GLA, resulted in the gain-of-function expression of GLA and octadecatetraenoic acid. The ubiquitous presence of octadecatetraenoic acid ($18{:}4^{\Delta 6,9,12,15}$) in GLA producing transgenic *Anabaena* provides additional insight into the C_{18} desaturation pathway. This unusual fatty acid, which is present normally in fish oils and in some plant species of the Boraginaceae family [12, 13] must result from the further desaturation of α-linolenic acid by a Δ^6-desaturase or desaturation of GLA by a Δ^{15}-desaturase. We further demonstrated that a 1.8 kb region of the *Synechocystis* genome contains both coding and promoter regions of the *Synechocystis* 2Δ^6-desaturase gene and is sufficient to produce GLA in *Anabaena* and *Synechococcus*, although in the latter case only when a second *Synechocystis* gene encoding Δ^{12}-desaturase is also present to generate linoleic acid.

The success of the gain-of-function approach described here, coupled with other molecular genetics tools now available in cyanobacteria, makes possible the identification of other cyanobacterial genes for which there is no selectable phenotype. Certainly, other genes involved in lipid metabolism are prime candidates; the triad of *Synechocystis*, *Anabaena* and *Synechococcus* provide an opportunity to isolate most genes involved in fatty acid metabolism in cyanobacteria. As a consequence, genes encoding desaturases for the

entire C_{18} fatty acid desaturation pathway soon will be available for study. This will facilitate analysis of the factors regulating levels of fatty acid desaturation and the role of these desaturation levels in overall cellular physiology, including chilling tolerance [25]. It is noteworthy that transgenic *Anabaena* and *Synechococcus* with altered levels of polyunsaturated fatty acids were similar to wild type in growth rate and morphology (data not shown) when grown under standard conditions; effects of lower temperatures were not examined. The availability of these genes also will allow detailed structure/function analysis of this class of desaturases and will provide further insight into the evolutionary constraints on protein structure and function.

Recently, transgenic tobacco plants were produced containing a chimeric cyanobacterial desaturase gene, comprised of the *Synechocystis* Δ^6-desaturase gene fused to sequences encoding a carrot extensin signal peptide [7] and an endoplasmic reticulum retention sequence (KDEL); expression of this chimeric gene was driven by a CaMV 35S promoter. These transgenic plants accumulated small but significant amounts of GLA (A.S. Reddy and T.L. Thomas, unpublished results). These results suggest that cyanobacterial genes involved in fatty acid metabolism can be used to generate transgenic plants with altered fatty acid compositions. These modifications could lead to improved nutritional characteristics or increased industrial value of seed oils or improved growth potential of crop plants. In addition, analysis of desaturase expression in a higher plant context may provide insight into the relative role of C_{18} desaturases in the chloroplast, presumably the more natural context of cyanobacterial desaturases, *vis à vis* the endoplasmic reticulum.

Acknowledgements

We are grateful to Dr James Golden and Dr Susan Golden for providing pDUCA7, pAM542 and pAM854 vectors, respectively, and access to their laboratories for the initial part of this work. We thank Dr James Golden and Dr Thomas McKnight for critical review of this manuscript. The work was supported by grants from Rhône-Poulenc Agrochimie, Lyon, France.

References

1. Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HM and Somerville CR: Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. Science 258: 1353–1355 (1992).
2. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K: Current Protocols in Molecular Biology. Greene Publishers/Wiley-Interscience, New York (1989).
3. Brenner RR: Regulatory function of Δ^6-desaturase - a key enzyme of polyunsaturated fatty acid synthesis. Adv Exp Med Biol 83: 85–101 (1976).
4. Buikema JW, Haselkorn R: Isolation and complementation of nitrogen fixation mutants of the cyanobacterium *Anabaena* sp. strain PCC 7120. J Bact 173: 1879–1885 (1991).
5. Burkardt H, Riess G, Puhler A: Relationship of group P1 plasmids revealed by heteroduplex experiments: RP1, RP4, R68 and RK2 are identical. J Gen Microbiol 114: 341–348 (1979).
6. Bustos SA, Golden SS: Expression of the *pSbDII* gene in *Synechococcus* sp. strain PCC 7942 requires sequences downstream of the transcription start site. J Bact 174: 7525–7533 (1991).
7. Chen J, Varner JE: An extracellular matrix protein in plants: characterization of a genomic clone for carrot extensin. EMBO J 4: 2145–2151 (1985).
8. Dahmer ML, Fleming PD, Collins GD, Hildebrand DF: A rapid, screening technique for determining the lipid composition of soybean seed. J Am Oil Chem Soc 66: 543–548 (1989).
9. Devereux J, Haeberli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 12: 387–395 (1984).
10. Elhai J, Wolk CP: Conjugal transfer of DNA to cyanobacteria. Meth Enzymol 167: 747–754 (1988).
11. Golden SS, Brusslan J, Haselkorn R: Genetic Engineering of the cyanobacterial chromosome. Meth Enzymol 153: 215–231 (1987).
12. Graig M, Bhatty MK: Naturally occurring all-*cis* 6,9,12,15-octadecatetraenoic acid in plant oils. J Am Oil Chem Soc 41: 209–211 (1964).
13. Gross ATH, Dorell DG: Seed and oil characteristics of *Onosmodium occidentale*. Can J Plant Sci 56: 659–664 (1976).
14. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–137 (1982).

15. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
16. Murata N: Low temperature effects on cyanobacterial membranes. J Bioenerget Biomembranes 21: 61–75 (1989).
17. Murata N, Nishida I: Lipids of blue green algae (cyanobacteria) In: Stumpf PK (ed) The Biochemistry of Plants, vol 9, pp. 315–347. Academic Press, Orlando, FL (1987).
18. Ohlrogge JB, Browse J, Somerville CR: The genetics of plant lipids. Biochim Biophys Acta 1082: 1–26 (1990).
19. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY: Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J Gen Microbiol 111: 1–61 (1979).
20. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
21. Shanklin J, Somerville C: Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. Proc Natl Acad Sci USA 88: 2510–2514 (1991).
22. Somerville CR, Browse J: Plant Lipids: metabolism, mutants and membranes. Science 252: 80–87 (1991).
23. Thiede MA, Ozols J, Strittmatter P: Construction and sequence of cDNA for rat liver stearyl coenzyme A desaturase. J Biol Chem 261: 13230–13235 (1986).
24. Thompson GA: Scherer DE, Foxall-Van Aken S, Kenny JW, Young HL, Shintani DK, Kridl JC, Knauf CV: Primary structures of the precursor and mature forms of stearoyl acyl carrier protein desaturase from safflower embryos and requirement of ferredoxin for enzyme activity. Proc Natl Acad Sci USA 88: 2578–2582 (1991).
25. Wada H, Gombos Z, Murata N: Enhancement of chilling tolerance of a cyanobacterium by genetic manipulation of fatty acid desaturation. Nature 347: 200–203 (1990).
26. Weete JD: Lipid Biochemistry of Fungi and Other Organisms. Plenum, New York (1980).

The Plant Journal (1992) 2(3), 397–404

TECHNICAL ADVANCE

Stringent repression and homogeneous de-repression by tetracycline of a modified CaMV 35S promoter in intact transgenic tobacco plants

Christiane Gatz*, Claus Frohberg and Regina Wendenburg
Institut für Genbiologische Forschung Berlin GmbH, Ihnestr. 63, 1 Berlin 33, Germany

Summary

A cauliflower mosaic virus (CaMV) 35S promoter derivative, which is tightly repressed by the Tn*10* encoded Tet repressor in a transient expression system as well as in transgenic plants has been constructed. After treatment of transgenic plants with tetracycline (Tc) the activity of the reporter enzyme β-glucuronidase (GUS) increased up to 500-fold in tissue culture as well as under greenhouse conditions. Efficient de-repression was achieved by Tc uptake through the roots as well as by Tc treatment of leaves of intact plants. As Tc is not very stable in the plants, this system can also be used for a transient expression of a transgene. This system provides a unique tool for regenerating transgenic plants carrying a repressed transgene and for efficiently de-repressing its activity by a specific inducer at any time point of further development.

Introduction

The ability to introduce foreign genes into the plant genome has provided the methodology to analyse the molecular mechanisms leading to co-ordinated expression of genes in transgenic plants (Schell, 1987). It also serves to express alien gene products or to modulate the expression of endogenous proteins (Sonnewald *et al.*, 1991). Especially the last option opens new avenues for analysing and understanding the contribution of a defined gene to the organism's phenotype (Berg, 1991). When using this approach, a regulated promoter is often desirable in order to induce expression at defined time points during development, or only in certain parts of a transgenic plant. In addition, a tightly repressed promoter is absolutely required if the expression of a certain gene product of interest interferes with the regeneration process.

A number of plant promoters regulated by light (Kuhlemeier *et al.*, 1987), heat (Ainley and Key, 1990), stress (Freeling and Bennett, 1985), or wounding (Keil *et al.*, 1989) are available for the controlled expression of a transgene. However, they suffer from the disadvantage that the inducing conditions influence a variety of responses in the plants. Therefore we have developed a tightly repressed, specifically de-repressible promoter by a suitable combination of bacterial control elements with a strong, normally constitutive plant promoter.

We have reported previously that the Tn*10* encoded Tet repressor can regulate the expression of a modified CaMV 35S promoter in transgenic tobacco plants (Gatz *et al.*, 1991). In principle, we generated a transgenic plant which constitutively synthesizes the bacterial repressor protein (tetR$^+$). Two binding sites for the Tet repressor, the 19 bp palindromic *tet* operators, were introduced downstream of the TATA-box of the normally constitutive CaMV 35S promoter. When stably integrated into the genome of the *tetR*$^+$ plant, only low levels of activity from this modified promoter were detected. An 80-fold increase in RNA levels was achieved after 0.5 h upon vacuum infiltration of single leaves with a buffer containing the inducer tetracyline (Tc, 0.1 mg l^{-1}), which prevents the repressor from binding to its operator sequences. Since then we have significantly improved the system by a further reduction of the expression in the uninduced stage using a different arrangement of the *tet* operators within the promoter. Moreover, we describe the effect of a variety of Tc application procedures, as well as the kinetics of induction in whole plants and the time course of the decline of the amount of GUS mRNA after omission of the Tc treatment.

Results and discussion

Combination of three tet *operators with the CaMV 35S promoter ('Triple-Op'-promoter)*

It has been proposed by Lin and Riggs that repression efficiencies increase with the number of operators within a promoter, if each copy by itself contributes to repression (Lin and Riggs, 1975). In two previous studies we have investigated the influence of single Tet repressor–operator

Received 20 September 1991; revised 6 December 1991.
*For correspondence (fax +49 30 830007 36).

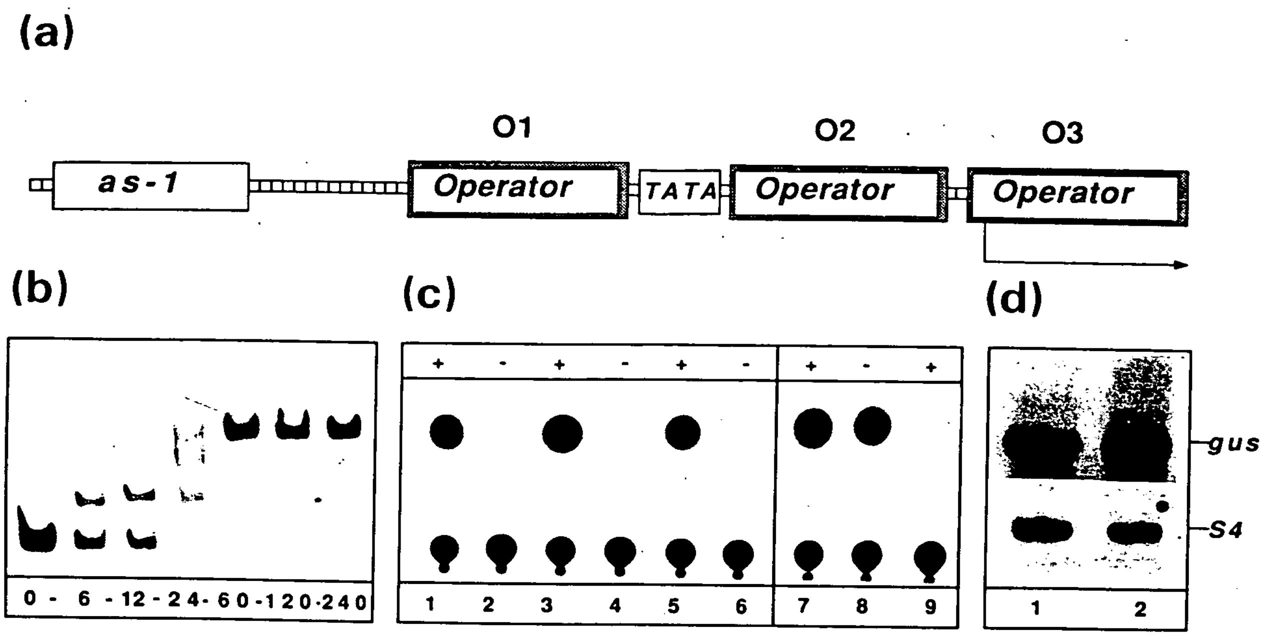


Figure 1. Schematic overview of gel shift analysis, transient expression analysis and Northern blot analysis of the 'Triple-Op' promoter.
(a) Region −90 to +17 of the 'Triple-Op'-promoter. Each square represents 1 bp. The sequence of activating sequence 1 (*as-1*, Lam *et al.*, 1989) is boxed, as well as the three operators (ACTCTATCACTGATAGAGT) and the TATA-box (TATATAA). The arrow marks the start site of transcription (Odell *et al.*, 1985).
(b) Mobility shift experiment demonstrating simultaneous occupation of the three operator sites within the 'Triple-Op'-promoter. One femtomole of a 350 bp fragment containing the complete 'Triple-Op'-promoter was incubated with increasing amounts of Tet repressor purified to homogeneity from *E. coli* (gift of Drs Altschmied and Hillen). Numbers below the lanes indicate the amount (fmol) of Tet repressor added.
(c) Cat activity in transiently transformed tobacco protoplasts which constitutively express the repressor gene (Gatz *et al.*, 1991). Ten micrograms of pTriple-Op-Cat (lanes 1–6) or pTET7 (lanes 7 and 9; Gatz and Quail, 1988) were introduced into tobacco protoplasts using polyethylene glycol mediated gene transfer and incubated with (+) or without (−) Tc. Cat activity shown in lane 9 represents background levels when protoplasts were transformed with herring sperm DNA.
(d) Northern blot analysis of Tc treated transgenic plants containing the chimeric 'Triple-Op'-promoter-GUS-gene (lane 1) or the wild-type CaMV 35S promoter-GUS-gene (lane 2). Rehybridization of the blot with the probe for the ribosomal gene *S4* was done to show that equal amounts of RNA were loaded. RNA from the three highest expressing plants of each transformation was combined for this analysis.

complexes in different positions on the expression of the CaMV 35S promoter (Frohberg *et al.*, 1991; Heins *et al.*, 1992). If located upstream of the TATA-box, efficient repression was only observed when the operator was located less than 3 bp away from the TATA-box. Downstream of the TATA-box the promoter was stringently repressed when the distance between the operator and the TATA-box was not more than 31 bp. In consideration of these data we constructed the so called 'Triple Op' promoter, which contained one operator (O1) 1 bp upstream of the TATA-box, a second operator (O2) 1 bp downstream of the TATA-box and a third operator (O3) 23 bp downstream of the TATA-box (Figure 1a). In Figure 1b we demonstrate that all three operators within this promoter fragment can simultaneously be occupied by the Tet repressor protein, though the spacing of 9 bp between O1 and O2 and the spacing of 2 bp between O2 and O3 is less than in the wild-type arrangement of 11 bp found between the two operator sites in the Tn*10* encoded regulatory region (Hillen *et al.*, 1984). With limiting amounts of Tet repressor four different bands can be observed in a mobility shift assay: free DNA, DNA bound to one repressor dimer, DNA bound to two repressor dimers and a fourth complex representing a fully saturated operator fragment. Next we analysed, in a transient expression system using chloramphenicol acetyl transferase (Cat) as a reporter enzyme (An, 1987), if the combination of the CaMV 35S promoter with three perfectly palindromic operator sequences affected promoter strength. As shown in Figure 1c no significant difference in gene expression was observed when comparing the 'Triple-Op'-promoter in the de-repressed stage with the wild-type CaMV 35S promoter. In the absence of the inducer, no detectable promoter activity was observed in *tetR*$^+$ protoplasts synthesizing the Tet repressor, indicating stringent repression. For the analysis of the promoter in transgenic plants, it was fused to the β-glucuronidase (GUS) gene (Jefferson *et al.*, 1987) and transferred to the genome of a *tetR*$^+$ plant using *Agrobacterium tumefaciens* mediated gene transfer. Leaves from 20 hygromycine resistant regenerated shoots were treated with Tc by vacuum infiltration (Gatz *et al.*, 1991) and GUS activity

was determined. All 20 plants showed no GUS activity before treatment with Tc and in every case a strong increase in GUS activity after Tc treatment. For control purposes a chimeric gene consisting of the wild-type CaMV 35S promoter (Covey and Hull, 1985) and the GUS gene was transferred to tobacco plants. Due to the variation of the expression levels of transgenes (Sanders *et al.*, 1987) we combined RNA from the three highest expressing plants from each transformation and subjected them to Northern blot analysis. As shown in Figure 1d, no difference in gene expression was observed, indicating that maximal wild-type promoter activities can be reached with the 'Triple Op'-promoter in the de-repressed stage.

Quantitation of the repression efficiency of the 'Triple-Op'-promoter in transgenic tobacco plants

At the level of GUS enzyme activity we consistently observed a 50-fold increase of expression after infiltration of single leaves with Tc, which is 10-fold more than we observed with one of our previous constructs which contains only two operators downstream of the TATA-box (Gatz *et al.*, 1991). Because all plants showed the same pattern of Tc dependent gene expression, we kept 10 of the highest expressing plants and randomly picked one of those for the various forms of analysis shown below. When analysing 30 μg total RNA on a Northern blot, we could not detect any GUS RNA in the repressed stage, with longer exposures yielding only background signals from cross-hybridizing ribosomal RNAs (data not shown). In order to quantitate the repression efficiency at the RNA level, we analysed poly(A)$^+$ RNA from repressed and de-repressed leaves of one of the 10 highest expressing transgenic plants. The amount of GUS mRNA in untreated and Tc-treated leaves was compared by using a dilution series of the signal obtained from Tc-treated leaves with mRNA from untransformed plants as a concentration standard. mRNA (800 ng) obtained from untreated and Tc treated leaves was loaded on a gel as well as a mixture (total amount: 800 ng) of poly(A)$^+$ RNA prepared from untransformed tobacco plants with 40, 26, 8, and 4 ng of the mRNA from Tc-treated plants. The signal obtained after rehybridization of the blot with a probe from the ribosomal gene *S4* (Devi *et al.*, 1989) was used to standardize the amount of mRNA loaded. Taking into account that about twofold more poly(A)$^+$ RNA was present in the lane containing mRNA from untreated leaves, we judged from the dilution series shown in Figure 2, that the activity of the 'Triple-Op'-promoter is repressed at least 100-fold at the RNA level.

De-repression of the 'Triple-Op' promoter in whole plants

With the repression efficiency being tight enough to observe significantly different levels of GUS activity in the repressed versus the de-repressed stage we started characterizing different modes of Tc application by doing *in-situ* stainings of whole plants with X-Gluc (Jefferson, 1987). First, we cultivated shoots of one transgenic plant on 2MS-medium supplemented with 1 mg l^{-1} Tc. Tobacco W38 forms roots without delay when grown in the presence of this amount of Tc. As shown in Figure 3a, no GUS activity was observed in the cutting that was grown without Tc even after an overnight incubation in X-Gluc. When grown on Tc-containing medium, however, dark blue staining representing high GUS activity was observed in the roots, the two lower leaves which had been in contact with the medium, and around the vascular tissue in some of the upper leaves. A leaf from a different cutting, which had only partly touched the medium showed staining in this region, and again around the vascular tissue. This result indicates that Tc is taken up through the roots and transported throughout the plant, and that it can also be taken up directly through the leaf. If we let a plant grow on 2MS medium without Tc and place one of the leaves between two separate blocks of agar containing Tc we observe after 6 h a local induction within this leaf. If we extend this treatment for 3 days we observe GUS activity in the lower and upper leaves as well as in the roots, which might indicate that Tc is transported through the phloem.

In order to achieve homogeneous distribution of the antibiotic throughout the plant we removed the lids of our tissue culture containers once a day for 15 min under sterile conditions thus enhancing transpiration. In addition we placed a piece of sterile cheesecloth between the lid and the container which also served to increase transpira-

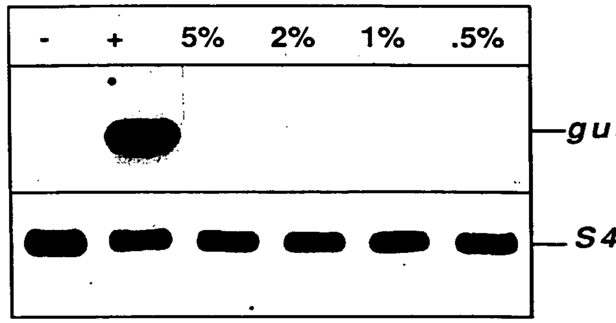


Figure 2. Northern blot analysis of poly(A)$^+$ RNA from a transgenic plant containing a chimeric repressor gene and a 'Triple-Op'-GUS-gene.
Poly(A)$^+$ RNA from an untreated and a Tc treated tobacco plant was analysed, as well as different amounts of poly(A)$^+$ RNA from the Tc treated plant mixed with mRNA from untransformed tobacco W38. Tc treatment was performed by vacuum infiltration of single leaves with 1 mg l^{-1} Tc and RNA was extracted after 2 days. Lane (−): 800 ng RNA from an untreated plant; lane (+): 800 ng RNA from a Tc-treated plant (TcRNA); lane (5%): 760 ng W38 RNA +40 ng TcRNA; lane (2%): 784 ng W38 RNA + 16 ng TcRNA; lane (1%): 792 ng W38 RNA + 8 ng TcRNA; lane (0.5%): 796 ng W38 RNA + 4 ng TcRNA. The blot was hybridized first with a GUS probe and afterwards with a *S4* probe.

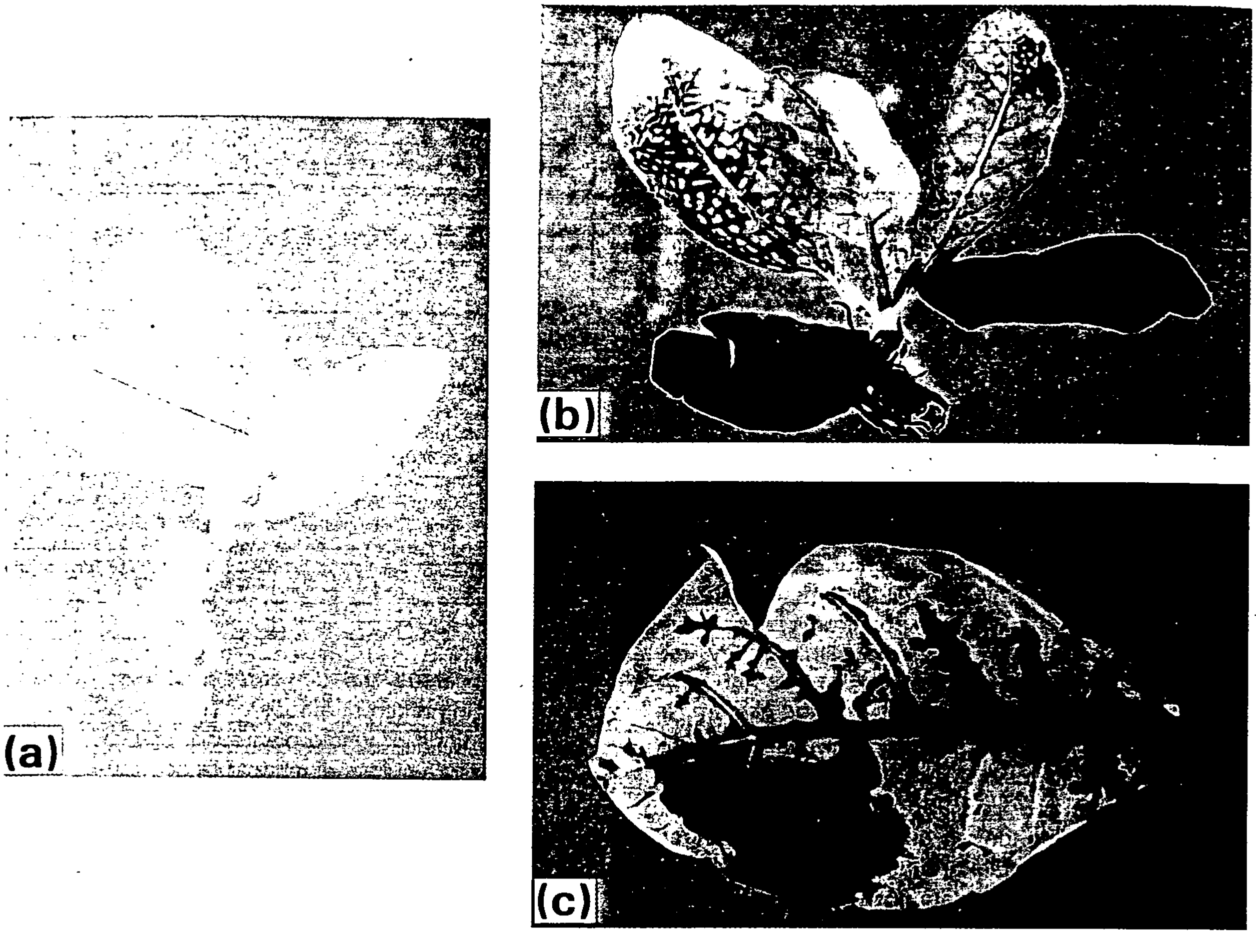


Figure 3. Localization of GUS enzyme activity after growing two cuttings of one transgenic plant on 2MS medium (a) which was supplemented with 1 mg l^{-1} Tc (b,c).
The dark blue staining represents high levels of GUS enzyme activity. The part of the leaf in (c) that was stained intensely blue had touched the medium. Plants had been grown on Tc containing medium for 2 weeks.

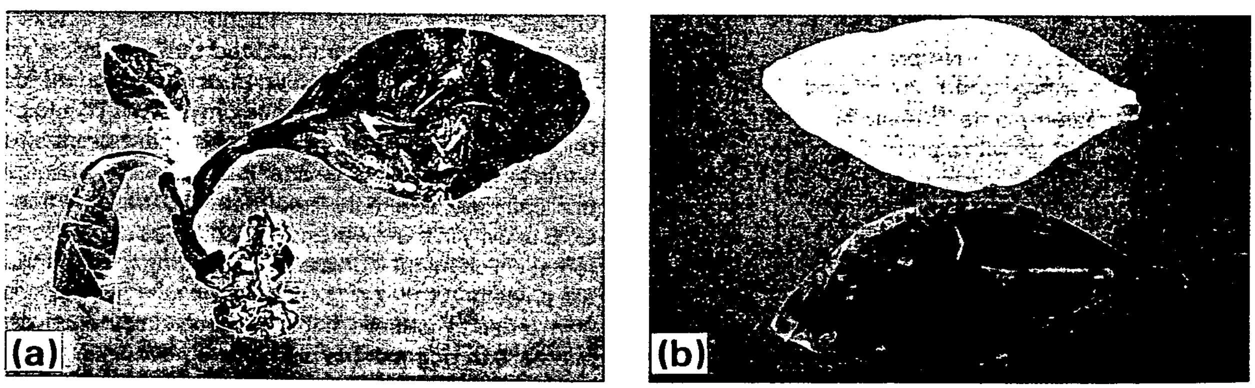


Figure 4. Localization of GUS enzyme activity.
(a) Localization of GUS enzyme activity after growing plantlets on vermiculite with Tc containing Hoagland buffer for 2 weeks. To enhance transpiration, the lid of the tissue culture container was removed under the hood for 15 min once a day. Every 3 days, Hoagland buffer containing fresh Tc was added.
(b) Direct comparison of an uninduced leaf with a leaf detached from the plant shown in (a).

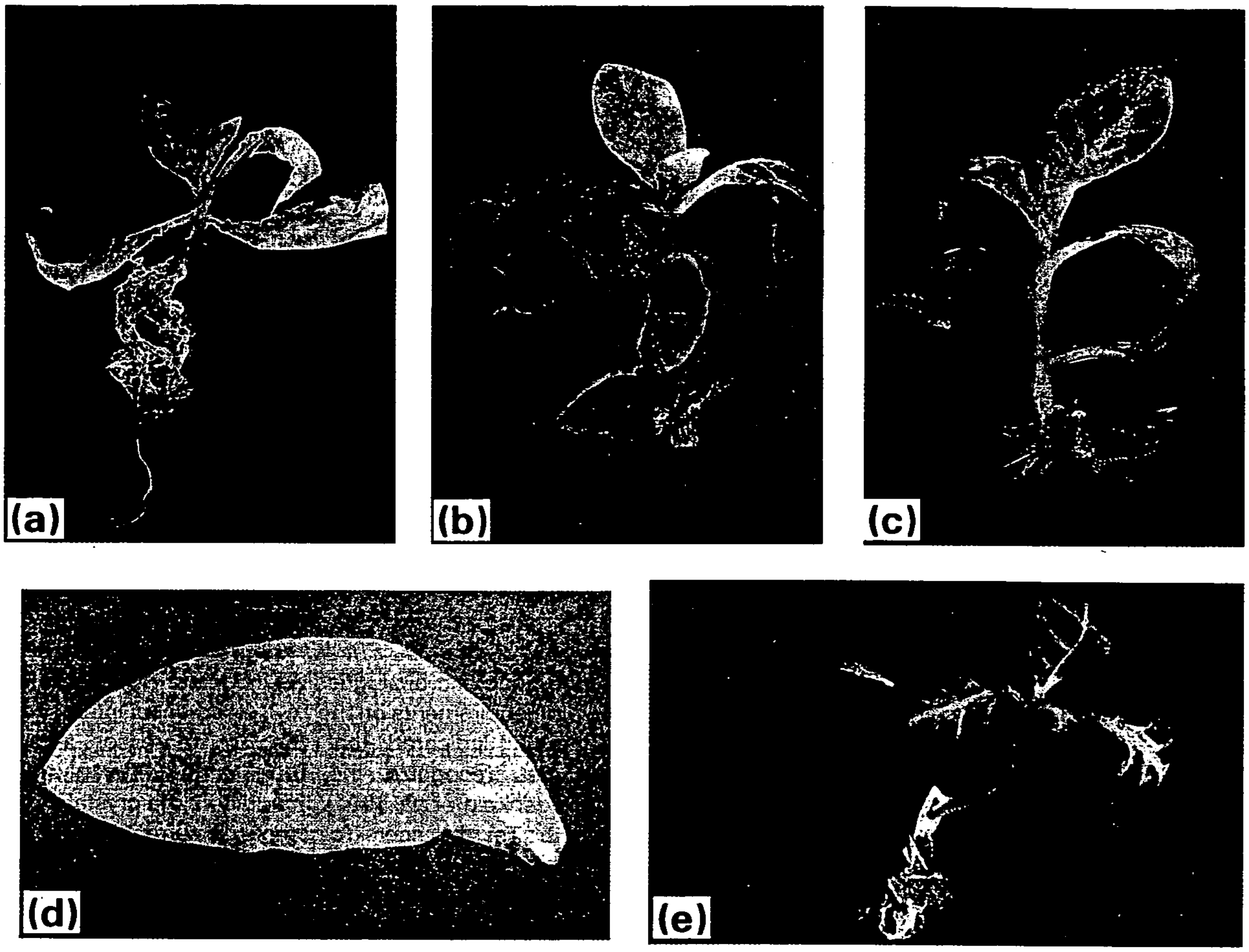


Figure 5. Localization of GUS enzyme activity after growing plantlets on vermiculite with Hoagland buffer.
Plantlets shown in (b) and (e) were submersed in a Tc containing buffer (1 mg l^{-1}) for 15 min once a day for 2 weeks. Plant in (a) is an independent cutting of the plant in (b), but grown on 2MS without Tc. The leaf shown in (d) is from an independent cutting of the plant in (e) that had been kept on 2MS without Tc. The plant in (c) was transformed with a chimeric wild-type CaMV 35S promoter–GUS construct.

tion in the growth chamber. As a second improvement we grew the plants on moist vermiculite, which had the advantage that fresh Tc could be added without cutting off the roots. As shown in Figure 4, homogeneous staining was observed throughout the plant, indicating sufficient distribution of the antibiotic. As determined by the fluorimetric GUS assay the gene was induced 500-fold under these conditions. This indicated that the 50-fold induction that had been measured 2 days after infiltration of single leaves might have been an underestimation because maximal levels of protein are not reached under these conditions.

As a third way of induction under tissue culture conditions we put whole plantlets into a breaker containing a buffer with 1 mg l^{-1} Tc. This type of Tc application also led to a GUS staining pattern that was indistinguishable from that of tobacco plants transformed with the wild-type 35S promoter fused to the GUS gene (Figure 5). Thus the light blue staining in the upper leaf is a property of the CaMV 35S promoter and not due to limited Tc uptake in younger leaves. In the long run, however, this mode of Tc application leads to some browning of the stem and the roots, so that we consider uptake through the roots as described above as more useful. Again, in this experiment, 500-fold induction at the level of GUS activity was observed.

In plants grown under greenhouse conditions, the promoter was de-repressed by applying the antibiotic through the roots (Figure 6). Plants suffered when sprayed with Tc in the presence of Saprogenate and uptake was poor when Saprogenate was omitted (data not shown).

Kinetics of de-repression and re-repression of the promoter

We have followed the kinetics of de-repression by taking samples from a Tc-treated plant and assaying them for

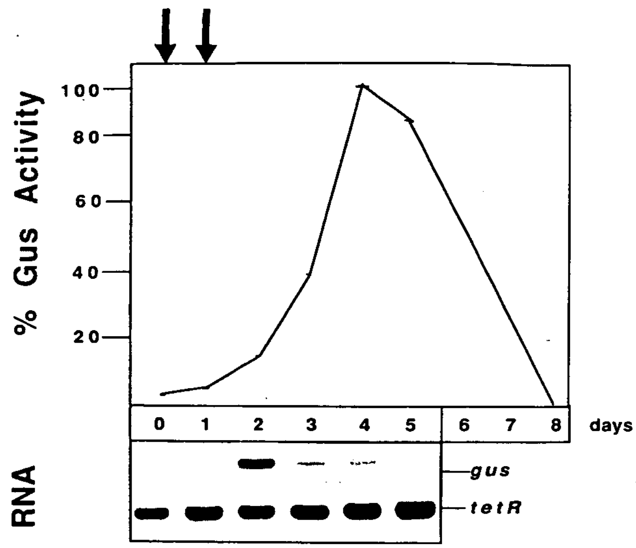


Figure 6. Time course of de-repression after Tc uptake through the roots in plants adapted to greenhouse conditions.
Arrows point to the 2 days where fresh Tc was added (1 mg l^{-1}). Each day a leaf was harvested for analysis of GUS enzyme activity and mRNA abundance.

GUS activity and RNA expression (Figure 6). As Tc uptake through roots had turned out to be the best way for de-repression of the promoter we used the following set-up for Tc treatment under greenhouse conditions. One of the transgenic plants, which contained six leaves at day 0, was cultivated with its roots hanging down into a beaker containing Hoagland buffer with fresh air being supplied through an aquarium pump. On day 0, Tc was added at a concentration of 1 mg l^{-1}, and the medium was exchanged only once on day 1 by a fresh batch of buffer, again containing 1 mg l^{-1} Tc. Each day a whole leaf was harvested for analysis of GUS activity and RNA extraction. Whereas maximal amounts of RNA were already detected on day 2, maximal levels of GUS enzyme activity (40-fold de-repression in this experiment) were observed on day 4. As long as detectable amounts of GUS RNA were present GUS activity continued to accumulate. We cannot clearly state at this moment if this—with respect to the RNA—delayed accumulation of the gene product is typical for the GUS reporter system or if it is caused by the untranslated leader, which contains at its 5′ end an almost complete palindromic operater sequence (Figure 1a). The kinetics of appearance of a gene product has to be specifically determined for each individual gene product, whereas RNA induction can be assumed to be maximal within a few days after addition of the antibiotic.

The antibiotic seems to become inactivated rather quickly in the plant because 2 days after the last addition of fresh Tc mRNA levels started to decrease and were indistinguishable from background levels after 4 days. This could be due to the light sensitivity of the antibiotic, because we do not see this effect when we incubate Tc treated leaves in the dark (Gatz *et al.*, 1991). One week after the last treatment with Tc, no GUS activity was detectable. This feature is extremely useful for the transient expression of a transgene, but for continuous de-repression fresh Tc has to be added at least every other day.

In conclusion, we have constructed a tightly repressed plant promoter that can be de-repressed with very low amounts of Tc. Using GUS as a reporter system we have demonstrated that homogeneous de-repression can be achieved by Tc uptake through the roots or through leaves. In terms of repression efficiency and de-repressibility the system seems to be suitable for the controlled expression of any transgene. It remains to be investigated, however, if plants carrying genes, which are lethal when expressed, can be regenerated using this system. Though Tc is applied at concentrations where we did not observe any phenotypic effect or a reduction of expression of the *S4* gene or the *tetR* gene under the control of the CaMV 35S promoter (Gatz *et al.*, 1991), we are going to develop non-antibiotic analogues as inducers. Studies on structural requirements on the Tc–Tet repressor interaction has already indicated that the Tet repressor and ribosomes recognize the drug in a different manner (Degenkolb *et al.*, 1991). In addition, the crystal structure of the Tet repressor –Tc complex is currently being solved (Parge *et al.*, 1984) so that detailed information on the synthesis of a non-antibiotic inducer should be available in the near future (W. Hillen, personal communication).

Experimental procedures

Plants, bacterial strains and media

Nicotiana tabacum L. was obtained through 'Vereinigte Saatzuchten' (Ebstorf, Germany). Plants in tissue culture were grown under a 16 h light/8 h dark regime on Murashige and Skoog medium (Murashige and Skoog, 1962) containing 2% sucrose (2MS) or on vemiculite with Hoagland buffer. *Escherichia coli* strains DH5α (Bethesda Research Laboratories, Gaithersburg, USA) and WH207/pRT241 (Wissmann *et al.*, 1986) were cultivated using standard techniques (Sambrook *et al.*, 1989). *Agrobacterium tumefaciens* strain C58C1 containing pGV2260 (Debleare *et al.*, 1984) was cultivated in YEB medium (Vervliet *et al.*, 1975).

Reagents

DNA restriction and modification enzymes were obtained from Boehringer Mannheim (Ingelheim, Germany) and New England Biolabs (Danvers, USA). Synthetic oligonucleotides were synthesized on an Applied Biosystems (Foster City, USA) DNA synthesizer (380A). Chemicals were obtained through Sigma Chemical Co. (St Louis, USA) or Merck (Darmstadt, Germany).

Recombinant DNA techniques

Standard procedures were used for recombinant DNA work (Sambrook *et al.*, 1989).

Constructs

First an oligonucleotide containing a *Spe*I- (−53), a *Sna*bI (−32), a *Stu*I- (−22), a *Xba*I (−16), a *Xho*I-(−3) and a *Bgl*II- (+2) site was insert d between the *Hga*I-site(−55) and the *Bgl*II site (+2) of pTET7 yielding pIGF107 using the same strategy as described previously (Gatz and Quail, 1988). Two complementary oligonucleotides with cohesive *Spe*I and *Bgl*II sites were synthesized: (oligonucleotide 1: CTAG-ACTCTATCAGTGATAGAGT-G-TATATAA-G-ACTCTATCAGTGATAGAGT-GA-ACTCTATCAGT GATACAGT-TAACGGTACCT, oligonucleotide 2: CTAGAGGTA CCGTTA-ACTCTATCACTGATAGAGT-TC-ACTCTATCACTGA TAGAGT-C-TTATATA-C-ACTCTATCACTGATAGAGT). This synthetic DNA fragment, which contained three operators, the CaMV 35S TATA-box as well as an *Hpa*I and an *Asp*718 site downstream of the third operator was inserted into pIGF107, cut with *Spe*I and *Xba*I, yielding pTriple-Op-Cat. Recombinant clones were detected using the repressor titration system described by Wissmann *et al.* (1986). This modified promoter was cloned as a *Sma*I–*Xba*I fragment in front of the GUS gene, using pAT3. The promoter in pAT3 was excised as an *Asp*718–*Xba*I fragment and was replaced by the 'Triple-Op'-promoter fragment after filling in the *Asp*718 site of pAT3. pAT3, which is a binary vector containing a hygromycin resistance gene, was used to transform a *tetR*$^+$ transgenic plant via *Agrobacterium tumefaciens* mediated gene transfer (Rosahl *et al.*, 1987).

Binding studies with purified Tet repressor

The 'Triple-Op'-promoter was excised as an *Eco*RI/*Bgl*II-fragment, purified from vector DNA using the 'Gene Clean' kit from Dianova (Hamburg) and end-labelled by filling in the protruding ends in the presence of [α-^{32}P]dATP using Klenow polymerase. Binding reaction and gel electrophoresis were carried out as described previously (Gatz *et al.*, 1991).

Transient expression in tobacco protoplasts

Isolation, transformation and chloramphenicol acetyl transferase assays were essentially as described by Frohberg *et al.* (1991).

Northern blot analysis

Total RNA from leaves was prepared according to Logemann *et al.* (1987). Poly(A)$^+$ RNA was prepared using the 'Dynabeads mRNA Purification Kit' from Dynal (Hamburg). Blotting and hybridization were carried out as described previously (Heyer and Gatz, 1991).

Assays for GUS activity

For the fluorometric GUS assay, explants were homogenized and incubated with the substrate 4-methylumbelliferyl-β-D-glucuronide at 37°C. Quantification of the fluorescence was done according to Jefferson (1987) and Jefferson *et al.* (1987). Protein concentrations were determined according to Bradford (1979).

For *in-vivo* staining, intact plant material was vacuum infitrated with 1 mM X-Gluc (5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid cyclohexylammonium) and incubated overnight at 37°C.

Tobacco transformation

Transformation of tobacco plants was carried out using the *Agrobacterium tumefaciens* leaf disc technique as described by Rosahl *et al.* (1987).

Application of tetracyline to the plants

For Tc application under axenic conditions, plants were either grown on 2 MS medium with 1 mg l^{-1} Tc, or on vermiculite in Hoagland buffer 1 mg l^{-1} Tc. Alternatively, plants were submersed once a day for 15 min in 1 mg l^{-1} Tc in 50 mM sodium citrate (pH 5.5). Aerial parts of plants adapted to greenhouse conditions were submersed once a day for 15 min in 1 mg l^{-1} Tc, 0.025% Saprogenate (Hoechst, Frankfurt) in 50 mM sodium citrate (pH 5.5). For Tc uptake through roots, plants were cultivated in a beaker containing Hoagland buffer and 1 mg l^{-1} Tc. Oxygen was supplied through an aquarium pump.

Acknowledgements

We thank Drs W. Hillen and L. Altschmied for purified Tet repressor, Dr Udo Schmitz and Hans-Peter Braun for the *S4* probe isolated from potato and Arnd Heyer for fruitful discussions. This work was supported by the Bundesministerium für Forschung und Technologie BCT 0389 'Molekulare und Zellbiologische Untersuchungen an höheren Pflanzen und Pilzen'.

References

Ainley, W.M. and Key, J.L. (1990) Development of a heat shock inducible expression cassette for plants: characterization of parameters for its use in transient expression assays. *Plant Mol. Biol.* **14,** 949–966.

An, G. (1987) Binary Ti vectors for plant transformation and promoter analysis. *Methods Enzymol.* **153,** 292–305.

Bradford, M.M. (1979) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein dye binding. *Anal. Biochem.* **72,** 248–254.

Berg, P. (1991) Reverse genetics: its origins and prospects. *Bio/Technol.* **9,** 342–344.

Covey, S.N. and Hull, R. (1985) Advances in cauliflower mosaic virus research. *Oxf. Surveys Plant Mol. Cell Biol.* **2,** 339–346.

Deblaere, R., Bytebier, B., De Greve, H., Debroeck, F., Schell, J., van Montagu, M. and Leemans, J. (1985) Efficient octopine Ti plasmid derived vectors for *Agrobacterium*-mediated gene transfer to plants. *Nucl. Acids Res.* **13,** 4777–4788.

Degenkolb, J., Takahashi, M., Ellestad, G. and Hillen, W. (1991) Structural requirements of tetracycline-Tet repressor interaction: Determination of equilibrium binding constants for tetracycline analogs with Tet repressor. *Antimicrob Agents Chemoth.* **35,** 1591–1592.

D vi, K.R., Chan, Y.-L. and Wool, I.G. (1989) The primary structure of rat ribosomal protein S4. *Biochem. Biophys. Acta,* **1008,** 258–262.

Freeling, M. and Bennett, D.C. (1985) Maize *adh1*. *Ann. Rev. Genet.* **19,** 297–332.

Frohberg, C., Heins, L. and Gatz, C. (1991) Characterization of the interaction of plant transcription factors using a bacterial repressor protein. *Proc. Natl Acad. Sci. USA*, **88**, 10470–10474.

Gatz, C. and Quail, P.H. (1988) Tn *10*-encoded Tet repressor can regulate an operator-containing plant promoter. *Proc. Natl Acad. Sci. USA*, **85**, 1394–1397.

Gatz, C., Kais r, A. and W nd nburg, R. (1991) Regulation of a modified CaMV 35S promoter by the Tn *10*-encoded Tet repressor in transgenic tobacco. *Mol. Gen. Genet.* **227**, 229–237.

Heins, L., Frohberg, C. and Gatz, C. (1992) The Tn *10*-encoded Tet repressor blocks early but not late steps of assembly of the RNA polymerase II initiation complex *in vivo*. *Mol. Gen. Genet.* in press.

Heyer, A. and Gatz, C. (1992) Isolation and characterization of a cDNA-clone coding for potato type A phytochrome. *Plant Mol. Biol.* **18**, 535–544.

Hillen, W., Schollmeier, K. and Gatz, C. (1984) Control of expression of the Tn *10*-encoded tetracyline resistance operon: II. Interaction of RNA polymerase and TET repressor with the *tet* operon regulatory region. *J. Mol. Biol.* **172**, 185–201.

J fferson, R.A. (1987) Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Mol. Biol. Rep.* **5**, 387–405.

J fferson, R.A., Kavanagh, R.H. and Bevan, M.W. (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907.

Keil, M., Sanchez-Serrano, J.J. and Willmitzer, L. (1989) Both wound-inducible and tuber specific expression are mediated by the promoter of a single member of the potato proteinase inhibitor II gene family. *EMBO J.* **8**, 1323–1330.

Kuhlemeier, C., Green, P.J. and Chua, N.-H. (1987) Regulation of gene expression in higher plants. *Ann. Rev. Plant. Physiol.* **38**, 221–257.

Lam, E., Benfey, P.N., Gilmartin, P.M., Fang, R.-X. and Chua, N.-H. (1989) Site-specific mutations alter *in vitro* factor binding and change promoter expression pattern in transgenic plants. *Proc. Natl Acad. Sci. USA*, **86**, 7890–7894.

Lin, S. and Riggs, A.D. (1975) The general affinity of *lac* repressor for *E. coli* DNA: implications for gene regulation in procaryotes and eucaryotes. *Cell*, **4**, 107–111.

Logemann, J., Schell, J. and Willmitzer, L. (1987) Improved method for the isolation of RNA from plant tissues. *Anal. Biochem.* **163**, 16–20.

Murashige, T. and Sk g, F. (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497.

Odell, G.T., Nagy, R. and Chua, N.-H. (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature*, **313**, 810–814.

Parge, H.E., Schneider, M., Hahn, U., Saenger, W., Altschmied, L. and Hillen, W. (1984) Crystallization of and preliminary X-ray diffraction data for TET-repressor and TET-repressor-tetracyline complex. *J. Mol. Biol.* **180**, 1189–1191.

Rosahl, S., Schmidt, R., Schell, J. and Willmitzer, L. (1987) Expression of a tuber-specific storage protein in transgenic tobacco plants: demonstration of an esterase activity. *EMBO J.* **6**, 1155–1159.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Sanders, P.R., Winter, J.A., Harnason, A.R., Rogers, S.G. and Fraley, R.T. (1987) Comparison of the cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants. *Nucl. Acids Res.* **15**, 1543–1558.

Schell, J. (1987) Transgenic plants as tools to study the molecular organisation of plant genes. *Science*, **237**, 1176–1183.

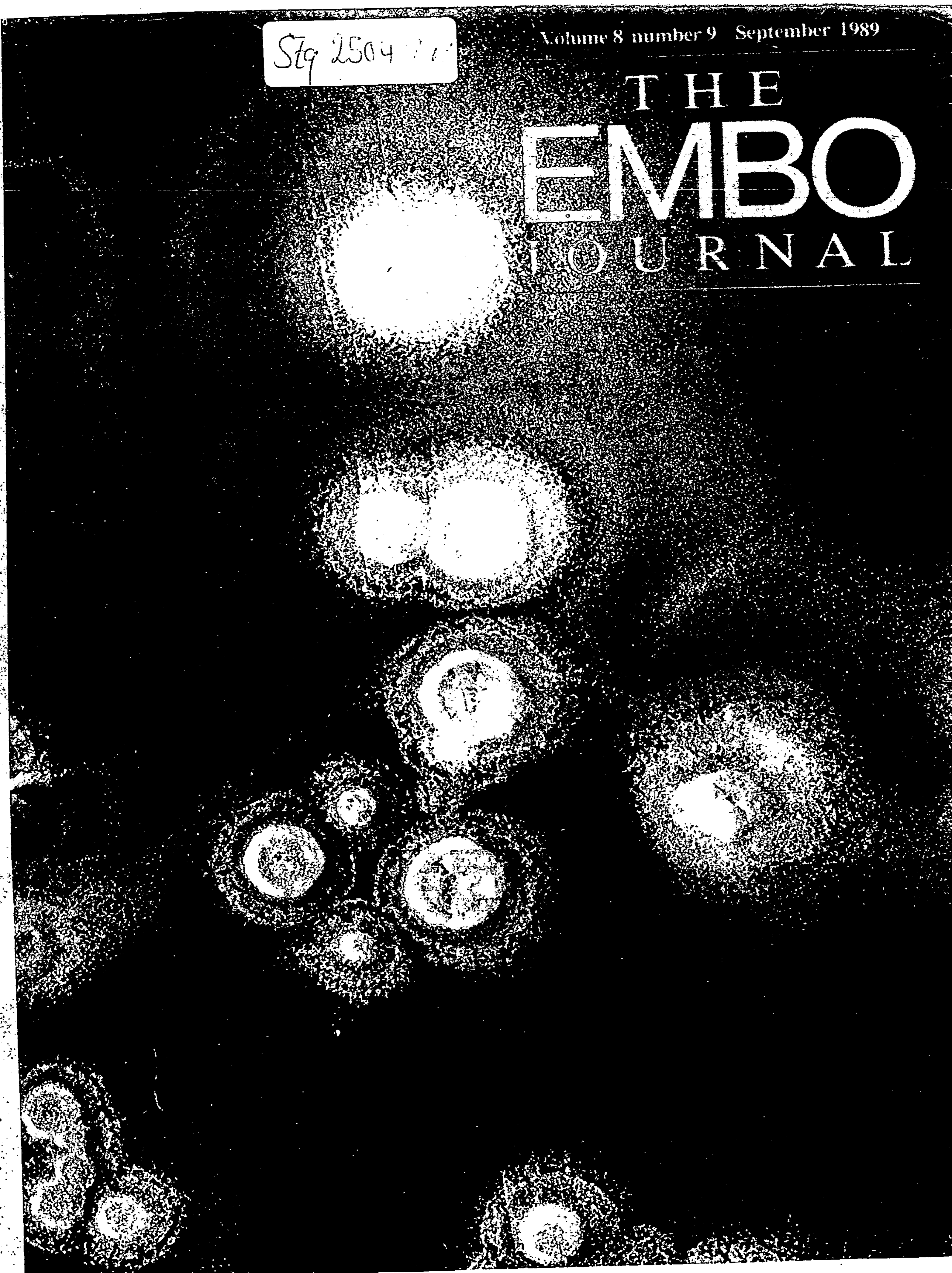
Sonnewald, U., Brauer, M., von Schaeven, A., Stitt, M. and Willmitzer, L. (1991) Transgenic tobacco plants expressing yeast-derived invertase in either the cytosol, vacuole or apoplast: a powerful tool for studying sucrose metabolism and sink/source interactions. *Plant J.* **1**, 95–106.

Vervliet, G., Holsters, M., Teuchy, H., Van Montagu, M. and Schell, J. (1975) Characterization of different plaque-forming and defective temperate phages in *Agrobacterium* strains. *J. Gen. Virol.* **26**, 33–48.

Wissmann, A., Meier, I., Wray, L.V. Jr, Geissendörfer, M. and Hillen, W. (1986) Tn *10* *tet* operator mutations affecting Tet repressor recognition. *Nucl. Acids Res.* **14**, 4253–4266.

Volume 8 number 9 September 1989

THE EMBO JOURNAL



The EMBO Journal vol.8 no.9 pp.2445–2451, 1989

Correlation of the expression of the nuclear photosynthetic gene ST-LS1 with the presence of chloroplasts

Jörg Stockhaus, Jeff Schell[1] and Lothar Willmitzer

Institut für Genbiologische Forschung Berlin GmbH, Ihnestrasse 63, D-1000 Berlin, and [1]Max-Planck-Institut für Züchtungsforschung, Egelspfad, D-5000 Köln 30, FRG

Communicated by J.Schell

A detailed analysis of the expression of a chimeric gene, consisting of the upstream region of the nuclear photosynthetic gene ST-LS1, encoding a component of the water-oxidizing complex of photosystem II, fused to the coding sequence of β-glucuronidase (GUS) as a reporter, is described. The expression of this chimeric gene at the cellular level was detected by histochemical methods and shows that the expression of this gene is correlated with the presence of chloroplasts. Interestingly, the GUS activity was not only detected in typical photosynthetic tissues, e.g. leaves and stems, but also in green roots containing chloroplasts. In contrast no activity was detected in neighbouring white root tissue which was devoid of chloroplasts. One can therefore separate the relative importance of the (morphological) differentiation steps responsible for the formation of tissues normally involved in photosynthesis, from the importance of the developmental stage (characterized by the presence of chloroplasts), for the expression of this nuclear photosynthetic gene. Our data strongly suggest that the developmental stage of the plastids is the primary determinant for the activity of this nuclear photosynthetic gene, although they do not yet allow the exclusion of the reverse type of control, i.e. control of the differentiation of the plastid by the expression of certain nuclear genes. A chimeric gene, consisting of the promoter of the 35S cauliflower mosaic virus (CaMV) gene and the GUS coding sequence, was used as a control throughout the experiments, confirming that the observed differential ST-LS1–GUS gene expression reflects the particular transcriptional regulation impacted on this gene by its *cis*-acting regulatory sequences.

Key words: cell-specific expression/chloroplast dependent expression/photosynthetic gene/35S promoter

Introduction

One important feature of eukaryotes is the fact that their genetic information is divided between two or, in the case of higher plants and algae, three different organelles—the nucleus, the mitochondria and the plastids. One important task for the cell is the coordination of the expression of genes present in these different cell compartments. This is of special importance in view of the fact that many plastidic and mitochondrial proteins are encoded by nuclear genes. Yeast mutants which are devoid of mitochondrial DNA, but nevertheless form organelles which structurally resemble mitochondria, are examples for the importance of the nuclear genome.

The photosynthetic apparatus of higher plants consists of several large protein complexes. As these complexes are encoded by both nuclear and plastidic genes, the plant cell therefore is faced with the problem of coordinating the expression of a large number of genes present in both compartments.

The molecular mechanisms which lead to this coordinated expression are unknown. In addition to light irradiation, which triggers the expression of several nuclear photosynthetic genes (Tobin and Silverthorne, 1985), the developmental stage of the cell is also important for their expression. In mature plants these genes are highly expressed in leaf mesophyll cells, whereas under natural growth conditions no expression is detectable in, for example, roots.

Several recent observations indicate that a 'plastidic factor' might be involved in the regulation of nuclear photosynthetic genes. It has been reported by Oelmüller and Mohr (1986) that the photo-oxidative damage of chloroplasts in mustard seedlings grown on a medium containing the herbicide Norfluorazon, leads to a severe reduction of the amount of translatable RNA, encoding the small subunit ribulose biphosphate carboxylase (RBCS) or the chlorophyll a/b binding protein (CAB). After a partial recovery of the chloroplasts, the amount of translatable mRNA increases again (Schuster *et al.*, 1988). Similar effects have been observed for the accumulation of CAB mRNA in carotenoid deficient tissue of maize seedlings where the carotenoid deficiency was due either to a mutation or to treatment with a herbicide (Mayfield and Taylor, 1984). Chlorophyll deficient maize seedlings, however, which contain plastids arrested in a developmental stage prior to chloroplast formation, accumulate normal levels of CAB mRNA (Mayfield and Taylor, 1984).

In the cases described above, the photo-oxidative damage of the chloroplasts did not affect the expression of several genes encoding cytoplasmic proteins (Reiß *et al.*, 1983; Mayfield and Taylor, 1984). These and other observations (Eckes *et al.*, 1985; Simpson *et al.*, 1986; Börner, 1986; Stockhaus *et al.*, 1987a) can be taken as indicative of a so-called 'plastidic factor', produced by the chloroplasts at a certain stage of development and which is essential for the expression of nuclear encoded chloroplastidic proteins. The observations summarized above are, however, hampered by the fact that all these data are based either on the use of inhibitors or of mutants leading to a photo-oxidative damage of the chloroplasts. With these experiments it is difficult to prove that the photo-oxidation will only influence the expression of the photosynthetic genes studied by the different authors and not result in any side effects. Furthermore these data are all based on the analysis of tissue homogenates. An analysis of the expression of these

genes at the single cell level would undoubtedly allow the establishment of a firmer correlation.

In order to overcome these limitations we decided to use an alternative approach which allows us to follow the expression of a well-characterized gene encoding a protein involved in photosynthesis at the single-cell level in intact 'wild-type' plants. Following this approach, we hoped to answer the question whether or not a close correlation of the activity of this gene with the developmental stage of the plastids could be established.

To this end we used β-glucuronidase (GUS) (Jefferson *et al.*, 1987) as a marker enzyme for the analysis of the cell specificity of ST-LS1 gene expression. ST-LS1 is a nuclear gene of potato, which was originally isolated by differential screening of cDNA libraries from leaves. This gene is expressed in a leaf/stem specific manner (Eckes *et al.*, 1985, 1986) and encodes a 10.8 kd protein associated with the oxygen evolving complex of photosystem II (Lautner *et al.*, 1988). For the experiments described below, a 1600 bp long upstream fragment of the ST-LS1 gene containing *cis*-elements sufficient both for high and specific expression of a chloramphenicol acetyltransferase (CAT) gene (Stockhaus *et al.*, 1987b) was fused to the GUS coding sequence. As a positive control we used a construct composed of the cauliflower mosaic virus (CaMV) 35S transcript (35S) promoter and the GUS coding sequence. These chimeric genes were introduced into potato (homologous system) and tobacco (heterologous system). The cell specificity of their expression was analysed in various organs and at various developmental stages at the cellular level. The observed correlation between ST-LS1−GUS gene expression and the presence of chloroplasts is discussed.

Results

Construction of ST-LS1−GUS and 35S−GUS genes and integration into the plant genome

The ST-LS1 gene upstream sequences from position −1600 to +11 (Stockhaus *et al.*, 1987b) were fused to the GUS coding sequence, followed by a polyadenylation signal of the T-DNA gene encoding the nopalinesynthase (abbreviated as ST-LS1−GUS) (see Figure 1A). A chimeric gene, consisting of the 35S promoter from position −526 to +4, GUS coding sequences and a polyadenylation signal derived from the CaMV 35S gene (see Figure 1B) (abbreviated as 35S−GUS) served as a positive control. By using the 35S promoter, we wanted to demonstrate that the absence of staining in some tissues (cf. below) observed for the ST-LS1−GUS construct is not due to artefacts caused by the lack of substrate in these cells, but rather to differential activity of the ST-LS1 promoter.

These constructs were inserted into the binary vector BIN19 (Bevan, 1984) and were used for the transformation of potato and tobacco plants via *Agrobacterium* strain pGV2260. The substrate 5-bromo-4-chloro-3-indolyl-glucuronide (X-Gluc) was used for the histochemical analysis of transgenic plants. This results in a blue staining of cells expressing the GUS enzyme (Jefferson *et al.*, 1987; Jefferson, 1987). In untransformed plants there was no GUS enzyme activity detectable in any of the tissues described here (data not shown). The data described below are based on the analysis of a number of independent transgenic potato and tobacco plants.

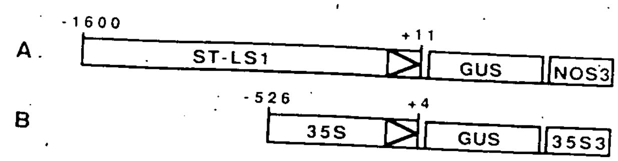


Fig. 1. Representation of the chimeric genes used for the analysis of the cell-specific expression patterns. **(A)** The ST-LS1 upstream sequences from position −1600 to +11 were fused to the GUS coding region and a polyadenylation signal of the nopalinesynthase gene (NOS3). **(B)** A chimeric gene consisting of the 35S upstream sequences from position −526 to +4, the GUS coding sequence (GUS) and the 35S polyadenylation signal (35S3) was used as a positive control.

ST-LS1−GUS and 35S−GUS gene expression in photosynthetic organs

The high level of steady-state ST-LS1 mRNA in leaves (Eckes *et al.*, 1985) indicated that this gene is strongly expressed in this organ. In order to identify the cells expressing the ST-LS1−GUS gene, the staining reaction was performed with transverse potato leaf sections. Very intensive staining of spongy mesophyll, palisade mesophyll and cells associated with the vascular bundles was observed (see Figure 2A). By electron microscopy it was demonstrated that parenchymatic cells which are associated with the vascular tissue contain chloroplasts. Parenchymatic cells of the middle rib of the leaf do not express the ST-LS1−GUS gene and do not contain chloroplasts.

This expression pattern is very similar to the one observed for the 35S−GUS gene which is also expressed to high levels in the palisade and spongy mesophyll cells as well as in parenchymatic cells associated with the vascular tissue (see Figure 2A).

The leaf epidermis is mainly composed of epidermal cells, stomata guard cells and trichomes. This tissue allows the comparison of the expression of the GUS fusions in photosynthetically inactive epidermal cells, containing rudimentary plastids, and photosynthetically active guard cells, which do contain chloroplasts. In contrast to the 35S−GUS gene, which is expressed both in epidermal and in guard cells (see Figure 2C), the ST-LS1−GUS gene expression is restricted to the chloroplast containing guard cells (see Figure 2B). This observation demonstrates that the substrate is not the limiting factor in epidermal cells, but that the staining pattern reflects the differential expression of the ST-LS1−GUS gene. It is furthermore remarkable that in tobacco trichomes both genes are expressed. In the small cells at the trichome tip which contain many chloroplasts the ST-LS1−GUS gene is however expressed to a higher level than the 35S−GUS gene (see Figure 2D and E).

In stem tissue of transgenic ST-LS1−GUS plants we detected low GUS enzyme activities in cells associated with the phloem tissue. In parenchymatic cells of either the pith or the stem-cortex respectively there was no detectable GUS enzyme activity (see Figure 2F). The 35S−GUS gene is highly expressed in parenchymatic cells of the phloem tissue (see Figure 2G). The analysis of longitudinal sections of the shoot apex confirmed the results obtained with the cross-sections. The ST-LS1−GUS gene expression is restricted mainly to the axillary buds and to parts of the apical meristem (see Figure 2H), whereas the 35S−GUS construct is highly expressed in the vascular tissue, axillary buds and in certain cells of the apical meristem (see Figure 2I).

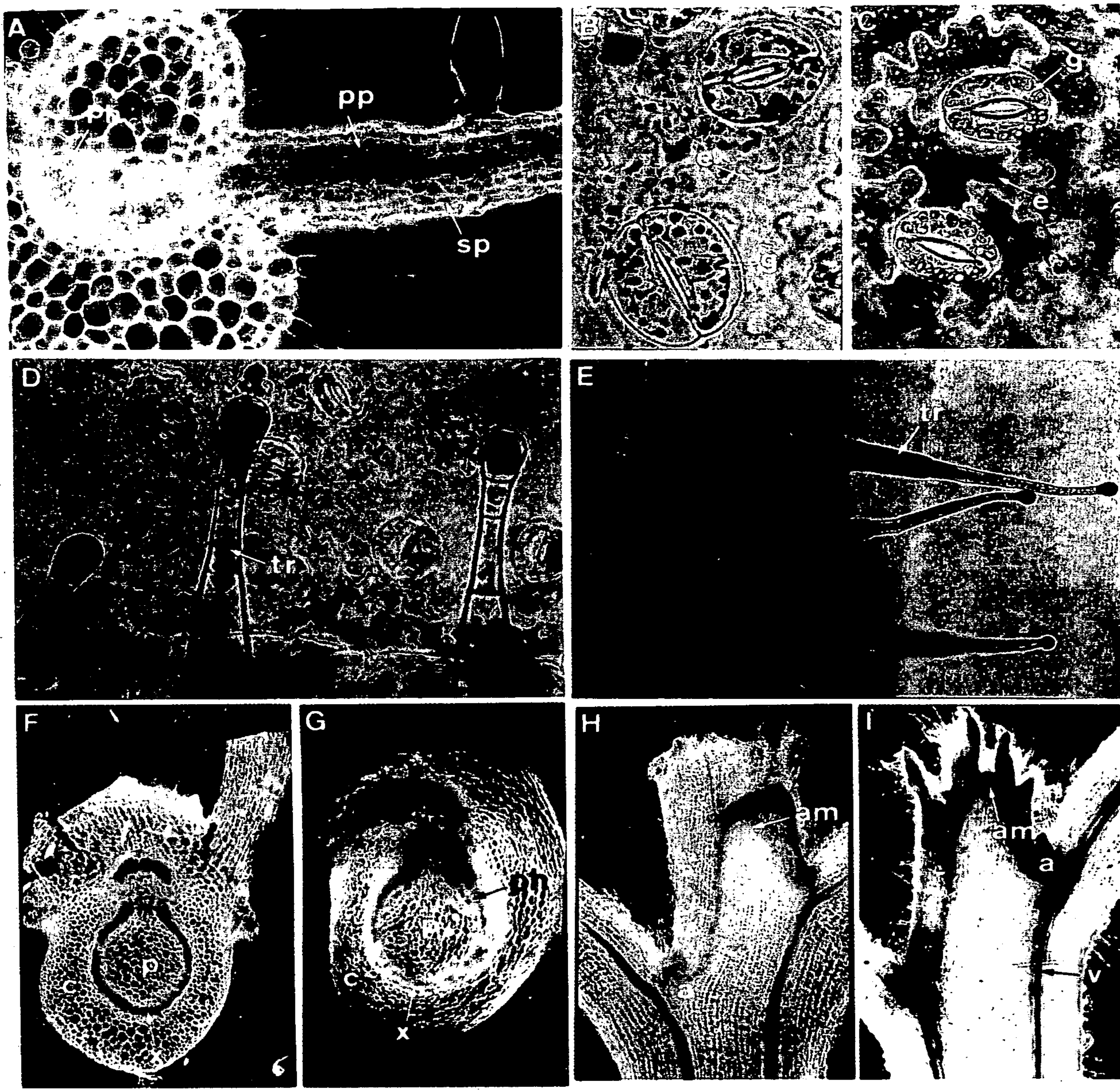


Fig. 2. Histochemical localization of the GUS enzyme activity in leaves and stem of tobacco plants transformed with the ST-LS1–GUS or the 35S–GUS gene. (**A**) Dark-field photograph of a transverse leaf section of a 35S–GUS plant. The dark blue staining represents high levels of GUS enzyme activity. Bright-field photographs of leaf epidermis of a ST-LS1–GUS plant (**B**) and a 35S–GUS plant (**C**); trichomes of a ST-LS1–GUS plant (**D**) and a 35S–GUS plant (**E**). Dark-field photographs of transverse stem sections of a ST-LS1–GUS plant (**F**) and a 35S–GUS plant (**G**); longitudinal sections of the shoot apex of a ST-LS1–GUS plant (**H**) and a 35S–GUS plant (**I**). a, axillary bud; am, apical meristem; c, cortex parenchyma; e, epidermal cell; g, guard cell; p, pith parenchyma; ph, phloem; pp, palisade parenchyma; sp, spongy parenchyma; tr, trichomes; v, vascular tissue; x, xylem.

Expression pattern in non-photosynthetic organs, e.g. roots and tubers of transgenic potato plants

In a second series of experiments we analysed the expression of the GUS fusions in organs characterized by the lack of chloroplasts under normal growth conditions.

The histochemical analysis of potato tuber cross-sections demonstrates that the ST-LS1–GUS gene is not expressed in tubers under normal growth conditions. In tubers exposed to white light for a few days, however, weak ST-LS1–GUS gene expression is detectable in rudimentary leaves of sprouting green buds (see Figure 3A) and the outer layer of chloroplast containing parenchymatic cortex cells (see Figure 3C). The 35S–GUS gene is expressed in parenchymatic cells associated with the vascular tissue in the pith (see Figure 3D) and in germinating buds (see Figure 3B). There was no expression detectable in the starch containing parenchymatic cells in the pith and in the periderm tissue of the tuber. Using transversal sections of roots of transgenic potato plants grown in soil we detected no ST-LS1–GUS gene expression (see Figure 3E), whereas the 35S–GUS

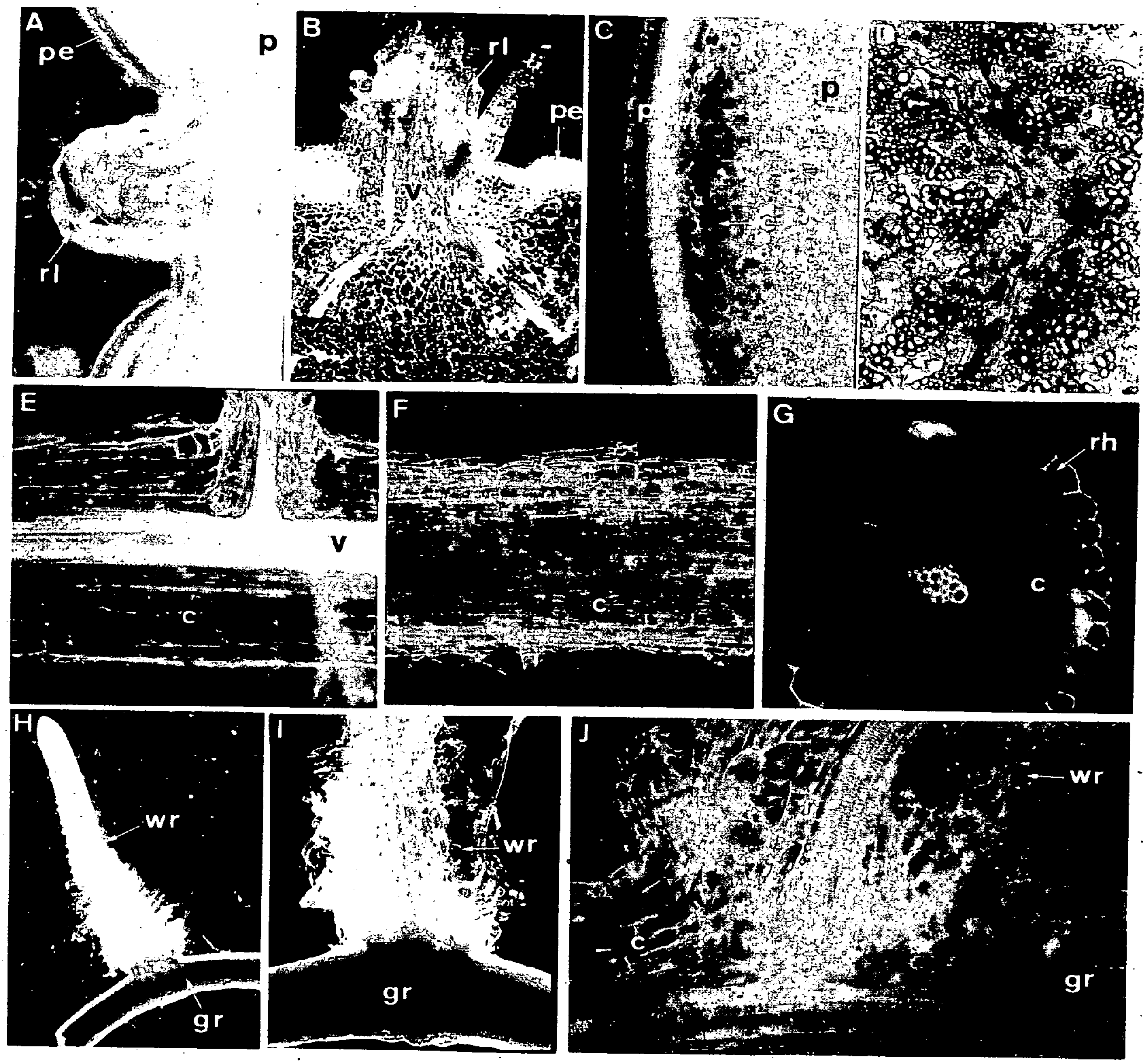


Fig. 3. Histochemical localization of GUS enzyme activity in roots and tubers of potato plants transformed with the ST-LS1−GUS or the 35S−GUS gene. Dark-field photographs of transverse sections of tubers of ST-LS1−GUS plants (**A**,**C**) and tubers of 35S−GUS plants (**B**,**D**); longitudinal section of a root of a ST-LS1−GUS plant grown in soil (**E**) and a 35S−GUS plant grown in soil (**F**); fluorescence photograph of a cross-section of a root of a potato plant grown in tissue culture (**G**). Dark-field photograph of an intact root of a tissue culture ST-LS1−GUS plant before (**H**) and after the GUS reaction (**I**); longitudinal section of a root of a ST-LS1−GUS plant grown in tissue culture (**J**). c, cortex parenchyma; gr, green root; p, pith parenchyma; pe, periderm; rh, rhizodermis; rl, rudimentary leaves; v, vascular tissue; wr, white root.

gene is highly expressed in the parenchymatic tissue of the root (see Figure 3F).

Roots of potato plants grown in tissue culture and which are therefore exposed to light do contain chloroplasts in parenchymatic cells (Eckes *et al.*, 1985) (see Figure 3G). The redifferentiation of these parenchymatic cells to chloroplast containing cells starts at a certain distance from the root tip. White side roots, growing out of older green roots (see Figure 3H), therefore represent a unique system allowing a direct comparison between ST-LS1−GUS gene expression in green roots, which contain chloroplasts, and young whitish roots which do not.

In parenchymatic cells containing chloroplasts a strong GUS enzyme activity is detectable (see Figure 3J), whereas there is no GUS enzyme activity detectable in the young outgrowing roots (see Figure 3I and J). In whitish roots of tobacco plants grown in tissue culture exposed to white light, we also observed chloroplasts by fluorescence microscopy, though their number is much lower. In these tobacco roots the ST-LS1−GUS gene is expressed, albeit at a rather low level (data not shown). In contrast to this highly differential expression of the ST-LS1−GUS gene in correlation to the

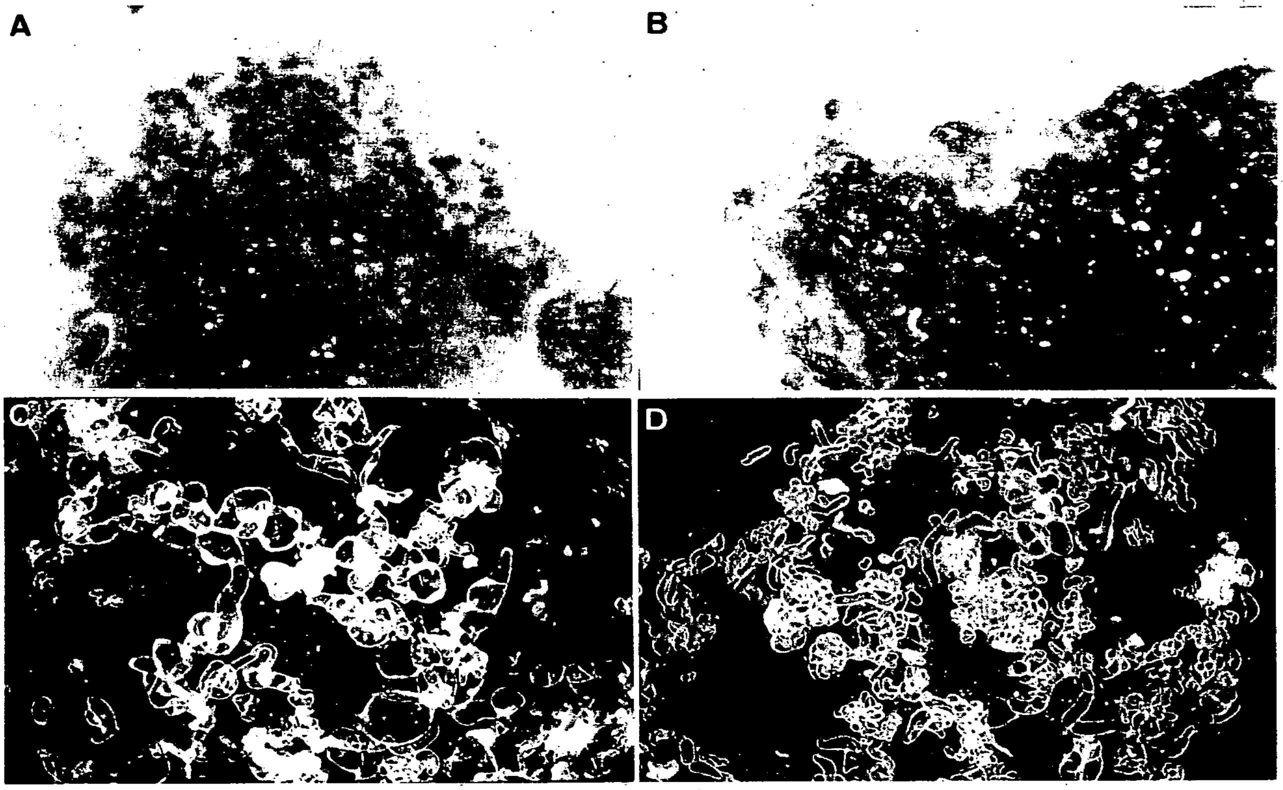


Fig. 4. Histochemical analysis of GUS enzyme activity in callus and suspension culture cells derived from potato plants transformed with the ST-LS1–GUS or the 35S–GUS gene. Bright-field photograph of a ST-LS1–GUS callus (**A**) and 35S–GUS callus (**B**). Dark-field photograph of ST-LS1–GUS suspension culture cells (**C**) and 35S–GUS suspension culture cells (**D**).

presence of chloroplasts, the 35S–GUS gene is expressed in white as well as in green parenchymatic root cells (data not shown).

Expression in potato callus and suspension culture cells

As a final step in our analysis, the expression pattern of both genes in undifferentiated callus and suspension culture cells was determined. A weak expression of the ST-LS1–GUS gene was detected in green callus cells (see Figure 4A). In callus cells representing a different developmental stage characterized by the lack of chloroplasts, no GUS activity was detected. The callus used for these experiments was derived from transgenic potato plants displaying high levels of GUS activity in leaves. The 35S–GUS gene is expressed to much higher levels in callus cells (see Figure 4B).

In tobacco as well as potato suspension culture cells grown under heterotrophic conditions and devoid of chloroplasts we again did not detect any ST-LS1–GUS gene expression (see Figure 4C). This contrasts with the high expression of the 35S–GUS gene in these cells (see Figure 4D).

Discussion

The photosynthetic apparatus localized in the chloroplasts of higher plants contains protein complexes which are encoded by the nuclear and the plastidic genome. In view of the central importance of the photosynthetic activity for the survival of the plant, it is obvious that the expression of the genes of both compartments must be interlinked and tightly controlled. Whereas post-transcriptional control appears to be especially important for the regulation of a number of plastidic genes (reviewed by Gruissem, 1989), the expression of nuclear photosynthetic genes appears to be regulated primarily at the transcriptional level. Light signal transducing systems in which phytochrome is involved play an essential role in this regulation (Tobin and Silverthorne, 1985). The coordinated expression of both nuclear and plastidic genes has, however, received less attention.

The data described in the Results point to a very strong correlation between the expression of a defined nuclear gene from potato (called ST-LS1), encoding a component of the water oxidizing complex of photosystem II, and the presence of chloroplasts. The three most striking examples for the correlation of the presence of chloroplasts with the expression of this nuclear photosynthetic gene are the data obtained for the leaf epidermis, root tissue and the potato tuber. In the epidermis of leaves, the ST-LS1–GUS gene is expressed in guard cells and trichomes which contain chloroplasts, whereas in epidermal cells which are devoid of chloroplasts there was no detectable ST-LS1–GUS gene expression. This result also indicates that, irrespective of the nature of the signal which is responsible for the induction of the ST-LS1 gene, it most likely has to be created within the cell itself and does not have any dominant influence on neighbouring cells. This signal therefore is unlikely to be able to diffuse or to be transported to other cells.

Our observation that the ST-LS1−GUS gene can be expressed in parenchymatic root and tuber cells, provided these tissues are made to contain green chloroplasts, represents an important finding with respect to the relative importance of the morphological differentiation of cells and the developmental stage of the plastids with regard to expression of the ST-LS1 gene. The observation that the ST-LS1−GUS gene is actively expressed in root and tuber cells containing chloroplasts whereas it is not expressed in neighbouring cells of the same type which are devoid of chloroplasts suggests that the presence of chloroplasts might be a prerequisite for the expression of the gene concerned. It should, however, be mentioned that they would also be compatible with an inverse type of control, i.e. control of the differentiation of the plastid by the expression of certain nuclear genes. These results also demonstrate that light—another factor often connected with the expression of photosynthetic nuclear genes—while essential, is not sufficient for induction of ST-LS1 gene expression. All further data described in the Results are in agreement with the main conclusion described above, i.e. the importance of the presence of chloroplasts for expression of the ST-LS1 gene. This result was obtained from the analysis in the homologous system (potato) as well as in the heterologous system (tobacco) for all tissues analysed.

The approach used in this study, i.e. the histochemical detection of β-glucuronidase activity from a chimeric gene transcriptionally driven by the promoter region of the ST-LS1 gene, was used for several reasons.

Firstly we wanted to know whether or not the postulated plastidary signal acts at the level of transcription. The chimeric gene used as a reporter consisted of regulatory sequences derived from a photosynthetic gene and of a coding sequence derived from a prokaryotic gene. We assumed that a prokaryotic mRNA would not be influenced markedly by plant specific post-transcriptional regulation mechanisms.

As outlined in the Introduction the importance of a plastidic factor for the expression of nuclear photosynthetic genes has been implied by several studies. These studies relied on the oxidative damage of chloroplasts by either the use of inhibitors of carotenoid biosynthesis or on the analysis of albino mutants. These previous data cannot with certainty exclude the possibility that the suppression of the activity of photosynthetic genes is due to a non-specific side effect of photo-oxidative damage. Our data, in contrast, were obtained in a 'wild-type' situation and in addition allowed us to monitor the expression on the cellular level.

It is important to examine whether or not the observed differential expression of the GUS enzyme is exclusively due to the specificity impacted by the ST-LS1 promoter. The expression of a chimeric 35S−GUS gene was therefore analysed in parallel and the expression patterns obtained for both genes were compared. This kind of analysis showed that the observed differential expression of the GUS gene results from ST-LS1 promoter activity and not, for example, from accessibility of the substrate or differences of GUS mRNA and protein stability.

Two other reports have to some extent described in a similar way the correlation between expression of another photosynthetic gene and the presence of chloroplasts. Using immunocytochemical methods, Aoyagi *et al.* (1988) showed that a chimeric gene consisting of the promoter of the nuclear photosynthetic small subunit RBCS gene fused to the coding sequence of the CAT gene was expressed in leaf and stem cells containing chloroplasts. A similar result was obtained by Jefferson *et al.* (1987) who demonstrated that treatment of stems with strong white light led to the formation of many chloroplasts in cortical parenchyma cells (chlorenchyma) and led to an increased level of expression of a chimeric gene consisting of a RBCS gene promoter fused to the GUS coding sequence. In these two cases the expression of the respective photosynthetic gene could not be separated from the formation of the typical photosynthetic tissues (leaves and stem). Nevertheless the observation that the *cis*-acting regulatory elements of different photosynthetic genes apparently led to the same kind of expression pattern as described for the ST-LS1 gene suggests that the hypothesized control of the expression of the ST-LS1 gene by the chloroplast could be a general phenomenon and might be relevant for a number of nuclear photosynthetic genes. The identification of tissues which, except for the difference of the developmental stage of their plastids, are very similar (tissue of green and white roots for example) will be very useful for the characterization of the signal(s) controlling the activity of nuclear photosynthetic genes.

Materials and methods

Recombinant DNA techniques

Standard procedures were used for recombinant DNA work (Maniatis *et al.*, 1982).

Transformation of tobacco and potato plants and tissue culture techniques

The chimeric genes were inserted in the vector BIN19 (Bevan, 1984) and introduced into the *Agrobacterium tumefaciens* strain pGV2260 (Deblaere *et al.*, 1985) by direct transformation according to Höfgen and Willmitzer (1988). In order to transfer the chimeric genes to tobacco cells, leaf discs of *Nicotiana tabacum* cv. SNN were infected with the respective *Agrobacterium* strain and subsequently regenerated (Horsch *et al.*, 1985). The transformation and regeneration of *Solanum tuberosum* cv. Desiree plants was performed as described by Rocha-Sosa *et al.* (1989).

Potato and tobacco callus was cultivated on MS medium (Murashige and Skoog, 1962) supplemented with 2% sucrose and 3 mg/l 2,4D (potato) or 1 mg/l 2,4D (tobacco) in a 16 h light/8 h dark rhythm. Suspension cultures were cultivated in liquid MS medium containing 2% sucrose and 1 mg/l 2,4D in continuous dim white light.

Histochemical localization

The histochemical reactions were performed as described by Jefferson (1987) using X-Gluc as substrate. For the sections of plant material a cryo-microtome was used. The staining reactions were performed with either unfixed cuttings or with cuttings fixed for 5−15 min in ice-cold 2% formaldehyde, 1 mM EDTA in 100 mM Na−phosphate (pH 7.0). The fixed cuttings were washed extensively before the staining reaction. The reaction times varied between 2 and 16 h.

Acknowledgements

We thank Carola Recknagel and Carola Mielchen for the tissue culture work and Regina Breitfeld and Beate Küsgen for taking care of the greenhouse plants. We are grateful to Richard Jefferson for providing the GUS coding region, to Reinhard Töpfer for the gift of the plasmid pRT102 and to Meike Köster for providing the plasmid pGUS. This work was supported by a grant from the Bundesministerium für Forschung und Technologie (BCT 0389: Molekulare und Zellbiologische Untersuchungen an höheren Pflanzen und Pilzen).

References

Aoyagi,K., Kuhlemeier,C. and Chua,N.-H. (1988) *Mol. Gen. Genet.*, **213**, 179−185.

Bevan,M. (1984) *Nucleic Acids Res.*, **12**, 8711–8721.
Börner,T. (1986) *Endocyt. C.R.*, **3**, 265–274.
Deblaere,R., Bytebier,B., De Greve,H., Deboeck,F., Schell,J., Van Montagu,M. and Leemans,J. (1985) *Nucleic Acids Res.*, **13**, 4777–4788.
Eckes,P., Schell,J. and Willmitzer,L. (1985) *Mol. Gen. Genet.*, **199**, 216–224.
Eckes,P., Rosahl,S., Schell,J. and Willmitzer,L. (1986) *Mol. Gen. Genet.*, **205**, 14–22.
Gruissem,W. (1989) *Cell*, **56**, 161–170.
Höfgen,R. and Willmitzer,L. (1988) *Nucleic Acids Res.*, **16**, 9877.
Horsch,R.B., Fry,F.E., Hoffmann,N.L., Eichholtz,D., Rogers,S.G. and Fraley,R.T. (1985) *Science*, **227**, 1229–1231.
Jefferson,R.A. (1987) *Plant Mol. Biol. Reporter*, **5**, 387–405.
Jefferson,R.A., Kavanagh,R.A. and Bevan,M.W. (1987) *EMBO J.*, **6**, 3901–3907.
Lautner,A., Klein,R., Ljungberg,U., Reiländer,H., Bartling,D., Andersson, B., Reinke,H., Beyreuther,K. and Herrmann,R.G. (1988) *J. Biol. Chem.*, **263**, 10077–10081.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Mayfield,S.P. and Taylor,W.C. (1984) *Eur. J. Biochem.*, **144**, 79–84.
Murashige,T. and Skoog,F. (1962) *Physiol. Planta.*, **15**, 473–497.
Oelmüller,R. and Mohr,H. (1986) *Planta*, **167**, 106–113.
Reiß,T., Bergfeld,R., Link,G., Thien,W. and Mohr,H. (1983) *Planta*, **159**, 518–528.
Rocha-Sosa,M., Sonnewald,U., Frommer,W., Stratmann,M., Schell,J. and Willmitzer,L. (1989) *EMBO J.*, **8**, 23–29.
Schuster,C., Oelmüller,R., Bergfeld,R. and Mohr,H. (1988) *Planta*, **174**, 289–297.
Simpson,J., Van Montagu,M. and Herrera-Estrella,L. (1986) *Science*, **233**, 34–38.
Stockhaus,J., Eckes,P., Blau,A., Schell,J. and Willmitzer,L. (1987a) *Nucleic Acids Res.*, **15**, 3479–3489.
Stockhaus,J., Eckes,P., Rocha-Sosa,M., Schell,J. and Willmitzer,L. (1987b) *Proc. Natl. Acad. Sci. USA*, **84**, 7943–7947.
Tobin,E.M. and Silverthorne,J. (1985) *Annu. Rev. Plant Physiol.*, **36**, 569–593.

Received on April 26, 1989; revised on June 8, 1989

Plant Molecular Biology

An International Journal on Fundamental Research and Genetic Engineering

Volume 22 / 1993

Kluwer Academic Publishers
Dordrecht / Boston / London

Plant Molecular Biology 22: 361–366, 1993.

Update section

Mini review

Chemical regulation of transgene expression in plants

E. R. Ward, J. A. Ryals and B. J. Miflin [1]*
Molecular Genetics Department, Agricultural Biotechnology Research Unit, Ciba-Geigy Corporation, P.O. Box 12257, Research Triangle Park, NC 27709, USA; [1] Research and Development, Seeds Division, Ciba-Geigy Limited, CH-4002 Basel, Switzerland (author for correspondence)*

Transgenic analysis has reached an advanced state in plants. In the ten years since transformation with a chimeric selectable marker was reported, the basic tools required to insert and express foreign genes have been developed. However, plant genetic engineers still lack important tools that are common in other systems; one that has only recently emerged is the ability to regulate expression of transgenes with exogenous chemical agents. This review will briefly cover the expanding literature on control of foreign gene expression in plants by application of synthetic compounds.

The goal of chemical gene induction systems is to provide the ability to manipulate levels of gene expression in order to understand better the functions of individual genes, and to facilitate the production of large amounts of a specific gene product. The basic concept underlying such schemes is isolation of a *cis*-acting sequence that operates as the key regulator of a gene with which it is naturally associated, followed by attaching the *cis*-acting element to a gene of interest. This results in expression of the engineered gene in a fashion similar to the natural, chemically regulated gene.

In other well-studied biological systems, the ability to alter gene expression by simple manipulation of the growth medium or addition of a chemical has found widespread use. Common systems include the lac operon in *Escherichia coli* [1], the *GAL 1, 4, 10* regulon in yeast [2], and the glucocorticoid receptor/response element in mammalian cells [3]. An important commercial use for chemical gene regulation is the production of recombinant proteins in fermentation settings (see e.g. [4]). Chemical control also provides the ability to study effects of ectopic expression of a specific promoter, spectacularly demonstrated in the 'super mouse' that arose from fusion of the metallothionein promoter to a rat growth hormone gene [5]. Thus, external regulation of gene expression serves the needs of both applied and basic science.

Combinations of *cis*-acting regulatory sequence and exogenous chemical regulator have been difficult to find in plants. An optimal combination of chemical inducer and target gene results in a tightly regulated system with very low uninduced expression that increases rapidly to high levels upon application of the inducer. The metabolic principles that underlie chemical gene regulation in microbes do not readily extrapolate to plants. For instance, simple inducers of catabolic processes (such as mono- and disaccharides) which are so useful in microbial systems are relatively useless in photoautotrophic organisms. Moreover, the possibilities of regulating the environment of auxotrophs in the field are much more limited than in fermentation systems. Transfer to special growth conditions for the sole purpose of gene induction will not be generally useful in agricultural settings, where conditions are optimized to maximize plant yield. Starvation for a particular nutrient or treatment with a chemical that produces phytotoxicity will be acceptable only in special situations. Natural plant metabolic signals and derivatives thereof are likely to be very

useful for regulating foreign gene expression; unfortunately, only a few such compounds and their target genes have been elucidated. Given the paucity of known natural regulators of plant gene expression, synthetic compounds have to date been more effectively used.

Two basic classes of chemical gene regulation can be distinguished: endogenous systems, which use regulatory signals from plant genes that respond to synthetic chemical treatment, and exogenous systems, which rely on elements from genes from other kingdoms, coupled with chemicals that have no effect on expression of native plant genes. Endogenous regulatory sequences are attractive in that they can be relatively easy to manipulate. For instance, the addition of a 5′ promoter element can be all that is needed to control a foreign coding sequence. On the downside, the use of endogenous plant regulatory sequences means that the native genes that are normally linked to these regulatory sequences are also induced upon addition of the chemical regulator. Thus, it is important to pick not only a potent inducer/regulatory sequence combination, but also an inducer of a class of genes that do not adversely affect the development of the plant. In addition, background levels of gene expression driven by the *cis*-acting sequence must not be overly sensitive to physiological changes in the plant that result from environmental fluctuation or other stresses. The basis for choosing an inducer that fills these criteria is largely empirical, and very few have been tested in intact plants to date. Four possibilities have been documented in the literature, and are reviewed briefly below.

Immunization compounds are chemicals that induce the systemic acquired resistance (SAR) response in plants. SAR is a broad resistance, effective against a variety of pathogens, that is induced by an initial pathogen infection [6] or chemical treatment. The inducing chemicals can be of natural origin, such as salicylic acid, or can be synthetic compounds, such as 2,6,-dichloroisonicotinic acid (INA) [7]. Treatment with a pathogen or an immunization compound induces the expression of at least nine sets of genes in tobacco, the best characterized species [8]. Different numbers and types of genes can be expressed in other plants [9, 10].

The promoter region of one tobacco gene, encoding pathogenesis-related (PR) protein 1a, has been demonstrated to confer chemically-inducible expression on the β-glucuronidase (GUS) reporter gene in laboratory settings [11–13]. We have shown that a PR-1a promoter/GUS fusion in transgenic tobacco behaves in the field as it does in the lab, reaching high expression levels after induction by either salicylic acid or INA (Fig. 1). Moreover, PR-1a promoter has recently been shown to drive chemically inducible expression of the insecticidal CryIA(b) protein of *Bacillus thuringiensis* [14]. This is the only example in the literature of chemical regulation of a potentially important transgenic agricultural trait. Plants expressing *B. thuringiensis* toxin specifically in response to an inducing stimulus may be

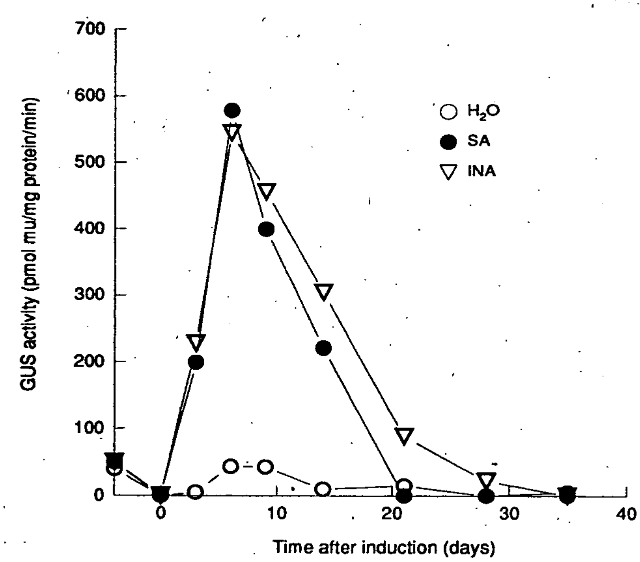


Fig. 1. Field performance of a PR-1a/GUS chimeric gene in transgenic tobacco. Six-week-old plants of a homozygous transgenic line containing a −903 PR-1a promoter fragment fused to GUS [13] were transplanted to the field 7 days before induction. Salicylic acid was applied at 50 mM; INA as a wettable powder formulation consisting of 25% active ingredient at 1 mg/ml. Each point represents the average of duplicate determinations from samples pooled on each day from three replicate field plots. SA, salicylic acid; mu, methylumbelliferone.

advantageous compared to constitutive expressers with respect to resistance management. Population genetic models in which *B. thuringiensis* toxin-resistant alleles are disadvantageous compared to wild-type (sensitive) alleles predict that resistance will spread more slowly when insect populations are given refuge from continual selection for resistance (reviewed in [15]). Thus, the ability to induce expression of the toxin would provide temporal refuge from selection, which should decrease the rate at which a population evolves toward resistance.

The PR-1a promoter/chemical inducer combination is also likely to find basic uses in plant biology. An INA-inducible PR-1 gene has been isolated from *Arabidopsis* (S. Uknes, E.R. Ward and J.A. Ryals, unpubl. data). In this intensively studied genetic model, the ability to control gene expression should be a valuable additional tool for studying the functions of individual genes of interest.

Safeners are chemicals known to induce the expression of enzymes involved in metabolism or detoxification of certain herbicides. The genes induced typically encode glutathione *S*-transferases [16, 17], cytochrome P-450 mixed-function oxygenases [18–20], or other proteins of unknown function. One group has reported the isolation of cDNAs that respond to safener treatment [21]. However, the level of induction for these genes is not as high as that seen for SAR-related genes induced by immunization compounds, which can be induced as much as 10000-fold over background [8, 10] (J.A. Ryals *et al.*, unpubl. data). No reports of chimeric constructions using safener-inducible elements have yet appeared, although such work is presumably in progress.

Genes involved in phenylpropanoid metabolism are known to be inducible by a variety of biotic and abiotic inducers [22]. Lamb and coworkers showed that a bean chalcone synthase promoter fused to the *E. coli uidA* gene in transgenic tobacco was inducible as much as 18-fold by pathogen infection, glucan elicitor from *Phytophthora megasperma*, and $HgCl_2$ [23]. Thus, in principle, regulatory sequences from the phenylpropanoid pathway could be coopted for other uses. Unfortunately, induction of the phenylpropanoid pathway may lead to accumulation of undesirable metabolites that are harmful to normal plant growth [22].

Work from the laboratory of C.A. Ryan over the past twenty years has focused on the wound induction of proteinase inhibitors in tomato and potato [24]. Experiments with transgenic tobacco showed that *cis*-acting sequences of a potato gene could confer wound inducibility on other genes, and that the relevant regulatory signals lay in the 3′ end of the gene [25]. Recently, outstanding progress has been made in elucidating the nature of the chemicals within the plant that signal wound induction systemically. The volatile lipid metabolite methyl jasmonate was found to induce expression of protease inhibitors in several plant species [26]. Despite the basic interest in this discovery, dosage and extent of coverage will probably be difficult to manipulate in systems using volatile compounds for artificial gene control, especially in field settings. More significantly, after an exhaustive search for the *in vivo* systemic inducer of protease inhibitors, an 18 amino acid peptide was found that induces gene expression in amounts as small as a few femtomoles [27]. The discovery of systemin, as this first peptide hormone from plants has been called, opens the door to a previously unexplored area of plant biochemistry. Presumably, expression of other inducible gene systems in plants may also be controlled by exceedingly potent peptides.

Exogenous regulators, which induce genes not occurring naturally in the plant, have the attractive feature of inducing only the introduced transgene. Unfortunately, adapting a gene control system from another organism means overcoming the formidable hurdles of ‘chemodynamics’. Specifically, the inducing compound must be (1) taken up efficiently by the plant, (2) moved systemically to the site of action, and (3) left in an active form by metabolic pathways that degrade or conjugate xenobiotics [28].

Schena *et al.* [29] recently showed that the mouse mammary tumor virus glucocorticoid receptor could confer inducibility on a truncated 35S promoter linked to several tandem copies of

the glucocorticoid response element in protoplasts. These experiments were carried out by cotransfection into tobacco protoplasts. The presence of the receptor gene caused dexamethasone-dependent induction of the GRE-driven reporter gene by as much as 150-fold. The absolute level of expression achieved was approximately 1/10 of that seen using a 35S promoter-driven reporter; presumably, higher levels of expression could be achieved by optimizing the conjunction of GRE to plant promoter elements. Similarly, the developmental and tissue specificity of the newly created inducible promoter could be varied by using individual elements from promoters specifically regulated in time or space during plant development. To date, however, the functioning of this otherwise attractive system has not been reported in stably transformed intact plants.

The Tn*10* tet repressor/operator is a prokaryotic control system that has been shown to function in plants [30]. Tobacco plants were first stably transformed with the *tetR* gene under 35S control. These plants expressed tet repressor at a level of ca. 0.01% of total cell protein. These TetR-expressing plants were transformed again, with a GUS reporter gene driven by a 35S promoter into which tet operator sequences had been integrated. Insertion of two tandem *tetO* sequences between the TATA box of the 35S promoter and the start of transcription conferred 50- to 80-fold repressibility on the GUS gene in the presence of tetracycline.

The definition of what constitutes an agricultural trait is changing quickly as genetic engineering of plants approaches its tenth anniversary. Chemical control has the capability to further expand the range of novel compounds that can be manufactured at commercially useful levels in plants. What are some potential uses of this technology?

One example is the use of plants as bioreactors to produce recombinant proteins. Van Montagu and coworkers showed that the neuropeptide Leu-enkephalin could be produced in the seed of transgenic *Arabidopsis* and *Brassica* by means of a translational fusion to a napin-like seed storage protein gene [31]. Conceivably, extremely high levels of such a peptide, in amounts that would draw deleteriously on the plant's N resources, could be synthesized under chemical control, followed quickly by harvest before significant starvation affected the crop.

Another recent example of a novel biosynthetic capacity conferred on plants through genetic engineering is the production of polyhydroxybutyrate [32]. This polyester thermoplastic is synthesized in three steps from acetyl-CoA by the bacterium *Alcaligenes eutrophus*. The first activity in the pathway, 3-ketothiolase, is found in plant cells. Somerville and coworkers introduced bacterial genes for the remaining two steps, each under the control of the 35S promoter, into *Arabidopsis thaliana*, creating two independent transgenic lines. An F1 hybrid of these lines accumulated PHB granules in the cytoplasm, vacuole, and nucleus. Expression of these genes was clearly harmful to the plant, as manifested by reduction in fresh weight between 20 and 45%. Thus, the ability to trigger a massive burst of PHB synthesis just prior to harvest might be an attractive strategy for its high-level production.

The chemical regulation of foreign genes will be especially powerful once homologous gene replacement becomes a routine technique in plant biology. Nearly all plant *cis*-regulatory sequences studied to date are exceedingly complex [33]. As a result, a defined fragment of a promoter linked to a different coding sequence and inserted into the genome in a random location hardly ever functions as well as it does in its native state. Once it becomes possible to swap a novel coding sequence into the milieu of a regulated locus, the foreign sequence will stand a much greater chance of being regulated like the native gene.

The types of traits that have been introduced into plants to date are relatively limited. While many of these traits will function adequately under constitutive expression, others will be useful only if regulated. As the biochemical bases for more complex plant processes are discovered, increasing numbers of transgenes are likely to be created that will be useful only if placed under exogenous control. The existence of easily used, highly controllable chemical gene regulation systems will

further drive the development of useful regulated phenotypes. Thus, the field of chemical induction of gene expression in plants is young indeed, and the traits that can be controlled using this burgeoning technology are limited only by the imaginations of investigators.

Acknowledgements

We thank Drs Mary-Dell Chilton, Bruce Lee and Scott Uknes for critically reading the manuscript. We also gratefully acknowledge Sandra Dincher, Shericca Williams, Patrick Kennedy, and Richard Lotstein for helping carry out the field test described in Fig. 1.

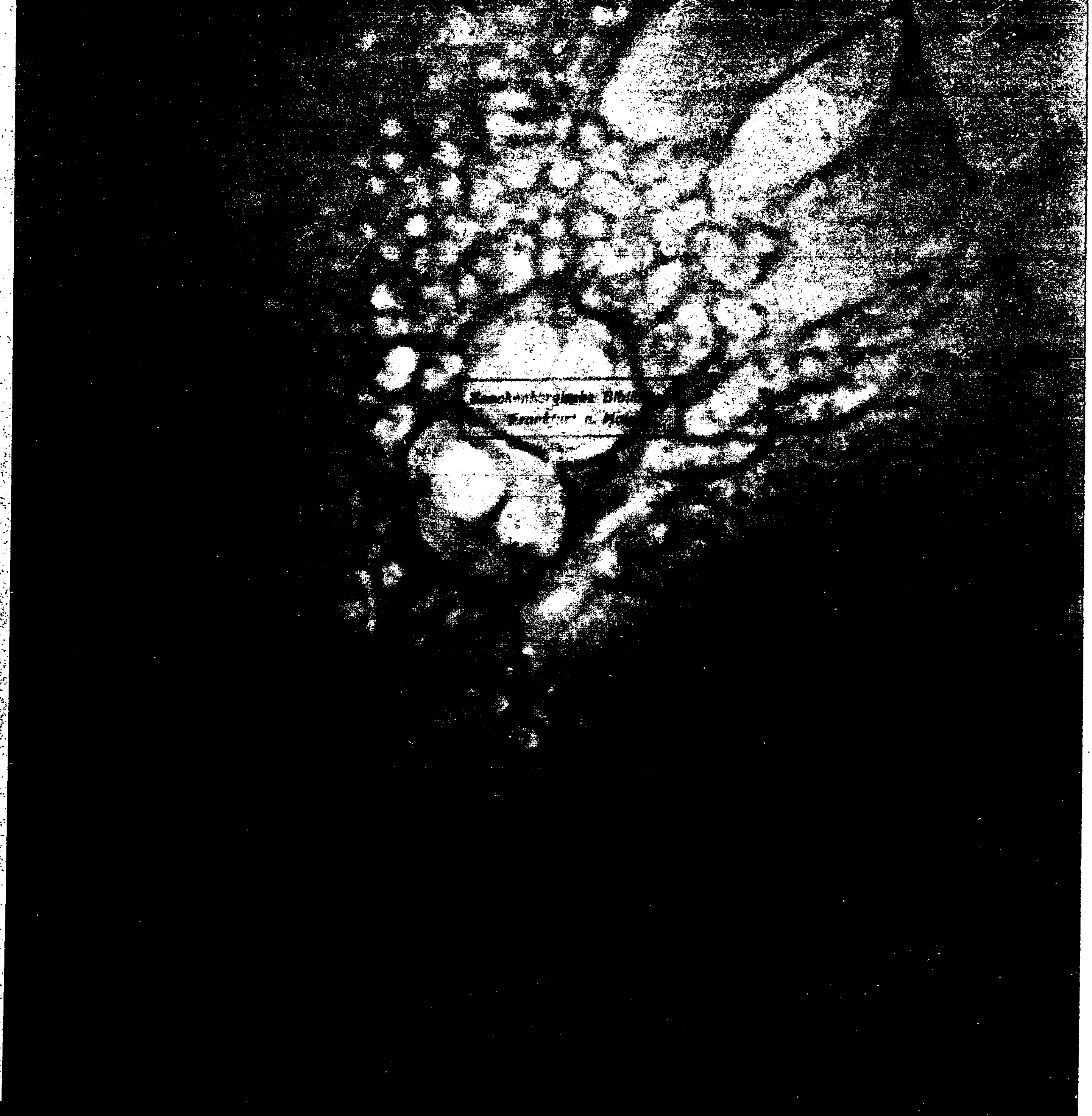
References

1. Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, Bolivar F, Boyer HW: Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science 198: 1056–1063 (1977).
2. Johnston M, Davis RW: Sequences that regulate the divergent *GAL1-GAL10* promoter in *Saccharomyces cerevisiae*. Mol Cell Biol 4: 1440–1448 (1984).
3. Chandler VL, Maler BA, Yamamoto KR: DNA sequences bound specifically by glucocorticoid receptor in vitro render a heterologous promoter hormone responsive in vivo. Cell 33: 489–499 (1983).
4. Sreekrishna K, Nelles L, Potenz R, Cruze J, Mazzaferro P, Fish W, Fuke M, Holden K, Phelps D, Wood P, Parker K: High-level expression, purification, and characterization of recombinant human tumor necrosis factor synthesized in the methylotrophic yeast *Pichia pastoris*. Biochemistry 28: 4117–4125 (1989).
5. Palmiter RD, Brinster RL, Hammer RE, Trumbauer ME, Rosenfeld MG, Birnberg NC, Evans RM: Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. Nature 300: 611–615 (1982).
6. Ross AF: Systemic acquired resistance induced by localized virus infections in plants. Virology 14: 340–358 (1961).
7. Métraux JP, Ahl Goy P, Staub T, Speich J, Steinemann A, Ryals J, Ward E: Induced resistance in cucumber in response to 2,6-dichloroisonicotinic acid and pathogens. In: Hennecke H, Verma DPS (eds) Advances in Molecular Genetics of Plant-Microbe Interactions, pp. 432–439. Kluwer, Dordrecht (1991).
8. Ward ER, Uknes SJ, Williams SC, Dincher SS, Wiederhold DL, Alexander DC, Ahl-Goy P, Métraux J-P, Ryals JA: Coordinate gene activity in response to agents that induce systemic acquired resistance. Plant Cell 3: 1085–1094 (1991).
9. Métraux JP, Burkhart W, Moyer M, Dincher S, Middlesteadt W, Williams S, Payne G, Carnes M, Ryals J: Isolation of a complementary DNA encoding a chitinase with structural homology to a bifunctional lysozyme/chitinase. Proc Natl Acad Sci USA 86: 896–900 (1989).
10. Uknes S, Mauch-Mani B, Moyer M, Potter S, Williams S, Dincher S, Chandler D, Slusarenko A, Ward E, Ryals J: Acquired resistance in *Arabidopsis*. Plant Cell 4: 645–656 (1992).
11. Ohshima M, Itoh H, Matsuoka M, Muramaki T, Ohashi Y: Analysis of stress-induced or salicylic acid-induced expression of the pathogenesis-related 1a protein gene in transgenic tobacco. Plant Cell 2: 95–106 (1990).
12. van de Rhee MD, van Kan JAL, González-Jaén MT, Bol JF: Analysis of regulatory elements involved in the induction of two tobacco genes by salicylate treatment and virus infection. Plant Cell 2: 357–366 (1990).
13. Uknes S, Dincher S, Friedrich L, Negrotto D, Williams S, Thompson-Taylor H, Potter S, Ward E, Ryals J: Regulation of pathogenesis-related protein-1a gene expression in tobacco. Plant Cell, in press (1993).
14. Williams S, Friedrich L, Dincher S, Carozzi N, Kessmann H, Ward E, Ryals J: Chemical regulation of Bacillus thuringiensis δ-endotoxin expression in transgenic plants. Bio/technology 10: 540–543 (1992).
15. McGaughey WH, Whalon ME: Managing resistance to *Bacillus thuringiensis* toxins. Science 258: 1451–1455 (1992).
16. Weigland RC, Shah DM, Mozer TJ, Harding EI, Diaz-Collier J, Saunders C, Jaworski EG, Tiemeier DC: Messenger RNA encoding a glutathione *S*-transferase responsible for herbicide tolerance in maize is induced in response to safener treatment. Plant Mol Biol 7: 235–243 (1986).
17. Timmerman KP: Molecular characterization of corn glutathione *S*-transferase isozymes involved in herbicide detoxification. Physiol Plant 77: 465–471 (1989).
18. Kreuz K, Gaudin J, Ebert E: Effects of the safeners CGA 154281, oxabetrinil and fenchlorim on uptake and degradation of metolachlor in corn (*Zea mays* L.) seedlings. Weed Res 29: 399–405 (1989).
19. Fonné-Pfister R, Kreuz K: Ring-methyl hydrozylation of chlortoluron by an inducible cytochrome P450-dependent enzyme from maize. Phytochemistry 29: 2793–2796 (1990).
20. Donaldson RP, Luster DG: Multiple forms of plant cytochromes P-450. Plant Physiol 96: 669–674 (1991).
21. Hershey HP, Stoner TD: Isolation and characterization of cDNA clones for RNA species induced by substituted benzenesulfonamides in corn. Plant Mol Biol 17: 679–690 (1991).
22. Hahlbrock K, Scheel D: Physiology and molecular biol-

ogy of phenylpropanoid metabolism. Annu Rev Plant Physiol Plant Mol Biol 40: 347–369 (1989).
23. Doerner PW, Stermer B, Schmid J, Dixon RA, Lamb CJ: Plant defense gene promoter-reporter gene fusions in transgenic plants: tools for identification of novel inducers. Bio/technology 8: 845–848 (1990).
24. Ryan CA: The search for the proteinase inhibitor-inducing factor, PIIF. Plant Mol Biol 19: 123–133 (1992).
25. Thornburg RW, An G, Cleveland TE, Johnson R, Ryan CA: Wound-inducible expression of a potato inhibitor II-chloramphenicol acetyltransferase gene fusion in transgenic plants. Proc Natl Acad Sci USA 84: 744–748 (1987).
26. Farmer EE, Ryan C: Interplant communication: airborne methyl jasmonate induces synthesis of proteinase inhibitors in plant leaves. Proc Natl Acad Sci USA 87: 7713–7716 (1990).
27. Pearce G, Strydom D, Johnson S, Ryan CA: A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. Science 253: 895–898 (1991).
28. Sandermann H: Plant metabolism of xenobiotics. Trends Biochem Sci 17: 82–84 (1992).
29. Schena M, Lloyd AM, Davis RW: A steroid-inducible gene expression system for plant cells. Proc Natl Acad Sci USA 88: 10421–10425 (1991).
30. Gatz C, Kaiser A, Wendenburg R: Regulation of a modified CaMV 35S promoter by the Tn*10*-encoded Tet repressor in transgenic tobacco. Mol Gen Genet 227: 229–237 (1991).
31. Vandekerckhove J, van Damme J, van Lijsebettens M, Botterman J, de Block M, Vandewiele M, de Clercq A, Leemans J, van Montagu M, Krebbers E: Enkephalins produced in transgenic plants using modified 2S seed storage proteins. Bio/technology 7: 929–932 (1989).
32. Poirer Y, Dennis DE, Klomparens K, Somerville C: Polyhydroxybutyrate, a biodegradable thermoplastic, produced in transgenic plants. Science 256: 520–523 (1992).
33. Benfey PN, Chua N-H: Regulated genes in transgenic plants. Science 244: 174–181 (1989).

Cell

Volume 21 Number 1 August 1980



Cell, Vol. 21, 285–294, August 1980, Copyright © 1980 by MIT

Nucleotide Sequence of Cauliflower Mosaic Virus DNA

A. Franck, H. Guilley, G. Jonard, K. Richards and L. Hirth
Laboratoire de Virologie
Institut de Biologie Moléculaire et Cellulaire du CNRS
15 rue Descartes
67084 Strasbourg Cédex, France

Summary

The complete nucleotide sequence (8024 nucleotides) of the circular double-stranded DNA of cauliflower mosaic virus has been established. The DNA molecule is known to possess three discrete single-stranded discontinuities, often referred to as "gaps," two in one strand and one in the other. The sequence data indicate that gap 1, the single discontinuity in the α strand, corresponds to the absence of no more than one or two nucleotides with respect to the complementary β strand. The two discontinuities in the β strand, however, are not authentic gaps since no nucleotides are missing, but are instead regions of sequence overlap: a short sequence (19 residues for gap 2, at least 2 residues for gap 3) at one terminus of each discontinuity, probably the 5′ terminus, is displaced from the double helix by an identical sequence at the other boundary of the discontinuity. Analysis of the distribution of nonsense codons in the DNA sequence is consistent with other evidence that only the α strand is transcribed. The coding region extends around the circular molecule from 4 map units of gap 1, the map origin, to map position 91, and consists of six long open reading frames. Our findings suggest, but do not prove, that the DNA sequence of the open reading frames is colinear with viral protein sequences. The cistron for the viral coat protein, which is probably synthesized in the form of a precursor, has been situated in coding region IV on the basis of its unusual amino acid composition.

Introduction

Cauliflower mosaic virus (CaMV) is the best characterized of the rather small number of plant viruses containing DNA rather than RNA as genetic material (for reviews see Hull, 1979a; Shepherd, 1979). CaMV DNA is double-stranded (Shepherd, Bruening and Wakeman, 1970) and has been estimated to be 7200–8000 bp long (Shepherd and Wakeman, 1971; Hull and Shepherd, 1977; Lebeuri r et al., 1978). Both lin ar and circular mol cules of similar contour l ngth may be found in CaMV DNA preparations (Shepherd and Wakeman, 1971; Russell et al., 1971; Civerolo and Lawson, 1978). Th circular form accounts for more than 90% of the material in fresh preparations and is the infectious entity (Hull and Shepherd, 1977; Volovitch, Drugeon and Yot, 1978); the linear DNA probably arises by adventitious breakage of circular molecules.

An unusual property of CaMV DNA is the existence of short discontinuities ("gaps") at well defined sites in one or the other strand of double-stranded circular molecules (Hull and Howell, 1978; Volovitch et al., 1978). Typically, there are two interruptions in one strand and one in the other (Volovitch et al., 1978; Hull, 1979b). We have chosen to designate gap 1, the single break in the transcribed α strand, as the zero point of our restriction map of circular CaMV DNA (Hohn et al., 1980), as there is evidence that RNA transcription begins near this point (Hull et al., 1979). The two gaps in the complementary β strand, gap 3 and gap 2, are located at 20 and 53 map units, respectively (Figure 2b). The positions of the three gaps are conserved in all CaMV isolates examined to date (Hull, 1979b) with the exception of CM184, which has undergone a small deletion in the region of gap 3 (Hull et al., 1979).

With the development of sophisticated techniques for constructing recombinant DNA molecules there has been a surge of interest in the possible use of DNA plant viruses such as CaMV as vectors for introducing foreign genes into plants. It is evident, however, that much will have to be learned about the molecular biology of these viruses and the way they interact with their hosts before such a plan can be tested. In this paper we report the complete nucleotide sequence of CaMV DNA (isolate Cabb B-S) and discuss those aspects of the sequence which shed light upon the organization of the CaMV genome.

Results and Discussion

Sequence Determination

A large number of Hinf I, Taq I, Mbo II and other double-stranded restriction fragments of CaMV DNA were prepared with ^{32}P-labeled 5′ termini by treatment with polynucleotide kinase in the presence of γ-^{32}P-ATP. After strand separation or secondary restriction to separate the labeled extremities, the sequence of the first 100–150 nucleotides in from each 5′ labeled terminus was determined by the limited chemical cleavage method of Maxam and Gilbert (1977). Enough data were collected to establish an unambiguous sequence for the entire genome, with over 75% of the molecule sequenced in both strands. (Details of this procedure are in Experimental Procedures.)

The complet sequence of CaMV DNA (isolatè Cabb B-S) is shown in Figure 1. Th sequ nce consists of 8024 bas s and numbering b gins with the first dG at the approximate 5′ boundary of gap 1. Only the sequence of the β strand, which has the sam polarity as viral mRNA (see b low), is presented.

Gap 1

GGTATCAGAG CCATGAATCG GTTTAAGACC AAAACTCAAG AGGGTAAAAC CTCACCAAAA TACGAAAGAG TTCTTAACTC 80

TAAAAATAAA AGATCTTTCA AGATCAAACA TAGTTCCCTC ACACCGGTGA CCGACAGGAT TACCACCGTA AGGTTTCAGA 160

ACAACATCGA AAGCGTTTAC GCCAACTTCG ACTCTCAACT CAAGTCGTCG TACGATGGTA GATCTAAAAA GATCAAGACT 240

CTAAGCCTTA AAAATCTTAG ATGTTACGAA GCCTTCCTCA GGAAGTACCT TCTGGAACAA TAAATCTCTC TGAGAATAGT 320

ACTCTATTGA GTATCCACAG GAAAAATAAC CTTCTGTGTT GAGATGGATT TGTATCCAGA AGAAAATACC CAAAGCGAGC 400

AATCGCAGAA TTCTGAAAAT AATATGCAAA TATTTAAATC AGAAAATTCG GATGGATTCT CCTCCGATCT AATGATCTCA 480

AACGATCAAT TAAAAAATAT CTCTAAAACC CAATTAACCT TGGAGAAAGA AAAGATATTT AAAATGCCTA ACGTTTTATC 560

TCAAGTTATG AAAAAAGCGT TTAGCAGGAA AAACGAGATT CTCTACTGCG TCTCGACAAA AGAATTATCA GTGGACATTC 640

ACGATGCCAC AGGTAAGGTA TATCTTCCCT TAATCACTAA GGAAGAGATA AATAAAAGAC TTTCCAGCTT AAAACCTGAA 720

GTCAGAAAGA CCATGTCCAT GGTTCATCTT GGAGCGGTCA AAATATTGCT TAAAGCTCAA TTTCGAAATG GGATTGATAC 800

CCCAATCAAA ATTGCTTTAA TCGATGATAG AATCAATTCT AGAAGAGATT GTCTTCTTGG TGCAGCCAAA GGTAATCTAG 880

CATACGGTAA GTTTATGTTT ACTGTATACC CTAAGTTTGG AATAAGCCTT AACACCCAAA GACTTAACCA AACCCTAAGC 960

CTTATTCATG ATTTTGAAAA TAAAAATCTT ATGAATAAAG GTGATAAAGT TATGACCATA ACCTATGTCG TAGGATATGC 1040

ATTAACTAAT AGTCATCATA GCATAGATTA TCAATCAAAT GCTACAATTG AACTAGAAGA CGTATTTCAA GAAATTGGAA 1120

ATGTCCAGCA ATCTGAGTTC TGTACAATAC AGAATGATGA ATGCAATTGG GCCATTGATA TAGCCCAAAA CAAAGCCTTA 1200

TTAGGAGCTA AAACCAAGAC TCAAATTGGT AATAACCTTC AAATAGGTAA CAGTGCTTCA TCCTCTAATA CTGAAAATGA 1280

ATTAGCTAGG GTAAGCCAGA ACATAGATCT TTTAAAGAAT AAATTAAAAG AAATCTGTGG AGAATAATAT GAGCATTACG 1360

GGACAACCGC ATGTTTATAA AAAAGATACT ATTATTAGAC TAAAACCATT GTCTCTTAAT AGTAATAATA GAAGTTATGT 1440

TTTTAGTTCC TCAAAAGGGA ACATTCAAAA TATAATTAAT CATCTTAACA ACCTCAATGA GATTGTAGGA AGAAGCTTAC 1520

TCGGAATATG GAAGATCAAC TCATACTTCG GATTAAGCAA AGACCCTTCG GAGTCCAAAT CAAAAAACCC GTCAGTTTTT 1600

Gap 3

AATACTGCAA AAACCATTTT TAAGAGTGGG GGGGTTGATT ACTCGAGCCA ACTAAAGGAA ATAAAATCCC TTTTAGAAGC 1680

TCAAAACACT AGAATAAAAA GTCTAGAAAA AGCAATTCAA TCCTTAGAAA ATAAGATTGA ACCAGAGCCC TTAACTAAAG 1760

AGGAAGTTAA AGAGCTAAAA GAATCGATTA ACTCGATCAA AGAAGGATTA AAGAATATTA TTGGCTAAAA TGGCTAATCT 1840

TAATCAGATC CAAAAAGAAG TCTCTGAAAT CCTCAGTGAC CAAAAATCCA TGAAAGCGGA TATAAAAGCT ATCTTAGAAT 1920

TATTAGGATC CCAAAATCCT ATTAAAGAAA GCTTAGAAAC CGTTGCAGCA AAAATCGTTA ATGACTTAAC CAAGCTCATC 2000

AATGATTGTC CTTGTAACAA AGAGATATTA GAAGCCTTAG GTACCCAACC TAAAGAGCAA CTAATAGAAC AACCTAAAGA 2080

AAAAGGTAAA GGCCTTAACT TAGGAAAATA CTCTTACCCC AATTACGGAG TAGGAAATGA AGAATTAGGA TCCTCTGGAA 2160

ACCCTAAAGC TTTAACCTGG CCCTTCAAAG CTCCAGCAGG ATGGCCGAAT CAATTTTAGA CAGAACCATT AATAGGTTTT 2240

GGTATAATCT GGGAGAAGAT TGTCTCTCAG AAAGTCAATT CGATCTTATG ATAAGATTGA TGGAAGAGTC CCTTGACGGG 2320

GACCAAATTA TTGATCTAAC CTCTCTACCT AGTGATAATT TGCAGGTTGA ACAGGTTATG ACAACTACCG AAGACTCAAT 2400

CTCGGAAGAA GAATCAGAAT TCCTTCTAGC AATAGGAGAA ACATCTGAAG AAGAAAGCGA TTCAGGAGAA GAACCTGAAT 2480

TCGAGCAAGT TCGAATGGAT CGAACAGGAG GAACGGAGAT TCCAAAAGAA GAAGATGGTG AAGGACCATC TAGATACAAT 2560

GAGAGAAAGA GAAAGACCCC GGAGGACCGG TACTTTCCAA CTCAACCAAA GACCATTCCA GGACAAAAGC AAACGTCTAT 2640

GGGAATGCTC AACATTGACT GCCAAACCAA TCGAAGAACT CTAATCGACG ACTGGGCAGC AGAAATCGGA TTGATAGTCA 2720

AGACCAATAG AGAAGACTAT CTCGATCCAG AAACAATTCT ACTCTTGATG GAACACAAAA CATCAGGAAT AGCCAAGGAG 2800

TTAATCCGAA ATACAAGATG GAACCGCACT ACCGGAGACA TCATAGAACA GGTGATCGAT GCGATGTACA CCATGTTCTT 2880

AGGACTAAAC TACTCCGACA ACAAAGTTGC TGAGAAGATT GACGAGCAAG AGAAGGCCAA GATCAGAATG ACCAAGCTCC 2960

AGCTCTGCGA CATCTGCTAC CTTGAGGAAT TTACATGTGA TTATGAAAAG AACATGTATA AGACAGAACT GGCGGATTTC 3040

CCACGATATA TCAACCAGTA CCTGTCAAAA ATCCCCATCA TTGGAGAAAA AGCGTTAACA CGCTTTAGGC ATGAAGCTAA 3120

CGGAACCAGC ATCTACAGTT TAGGTTTCGC GGCAAAGATA GTCAAAGAAG AACTATCTAA AATCTGCGAC TTATCCAAGA 3200

AGCAGAAGAA GTTGAAGAAA TTCAACAAGA AGTGTTGTAG CATCGGAGAA GCTTCAACAG AATATGGATG CAAGAAGACA 3280

TCCACAAAGA AGTATCACAA GAAGCGATAC AAGAAAAAAT ATAAGGCTTA CAAACCTTAT AAGAAGAAAA AGAAGTTCCG 3360

ATCAGGAAAA TACTTCAAGC CCAAAGAAAA GAAGGGCTCA AAGCAAAAGT ATTGCCCAAA AGGCAAGAAA GATTGCAGAT 3440

GTTGGATCTG CAACATTGAA GGCCATTACG CCAACGAATG TCCTAATCGA CAAAGCTCGG AGAAGGCTCA CATCCTTCAA 3520

CAAGCAGAAA AATTGGGTCT CCAGCCCATT GAAGAACCCT ATGAAGGAGT TCAAGAAGTA TTCATTCTAG AATACAAAGA 3600

AGAGGAAGAA GAAACCTCTA CAGAAGAAAG TGATGGATCA TCTACTTCTG AAGACTCAGA CTCAGACTGA GCAGGTGATG 3680

AACGTCACCA ATCCCAATTC GATCTACATC AAGGGAAGAC TCTACTTCAA GGGATACAAG AAGATAGAAC TTCACTGTTT 3760

CGTAGACACG GGAGCAAGCC TATGCATAGC ATCCAAGTTC GTCATACCAG AAGAACATTG GGTCAATGCA GAAAGACCAA 3840

TTATGGTCAA AATAGCAGAT GGAAGCTCAA TCACCATCAG CAAAGTCTGC AAAGACATAG ACTTGATCAT AGCCGGCGAG 3920

ATATTCAGAA TTCCCACCGT CTATCAGCAA GAAAGTGGCA TCGATTTCAT TATCGGCAAC AACTTCTGTC AGCTGTATGA 4000

ACCATTCATA CAGTTTACGG ATAGAGTTAT CTTCACAAAG AACAAGTCTT ATCCTGTTCA TATTGCGAAG CTAACCAGAG 4080

CAGTGCGAGT AGGCACCGAA GGATTTCTTG AATCAATGAA GAAACGTTCA AAAACTCAAC AACCAGAGCC AGTGAACATT 4160

Gap 2

TCTACAAACA AGATAGAAAA TCCACTAGAA GAAATTGCTA TTCTTTCAGA GGGGAGGAGG TTATCAGAAG AAAAACTCTT 4240

TATCACTCAA CAAAGAATGC AAAAAATCGA AGAACTACTT GAGAAAGTAT GTTCAGAAAA TCCATTAGAT CCTAACAAGA 4320

CTAAGCAATG GATGAAAGCT TCTATCAAGC TCAGCGACCC AAGCAAAGCT ATCAAGGTTA AACCCATGAA GTATAGCCCA 4400

ATGGATCGCG AAGAATTTGA CAAGCAAATC AAAGAATTAC TGGACCTAAA AGTCATCAAG CCCAGTAAAA GCCCTCACAT 4480

GGCACCAGCC TTCTTGGTCA ACAATGAAGC CGAGAAGCGA AGAGGAAAGA AACGTATGGT AGTCAACTAC AAAGCTATGA 4560

ACAAAGCTAC TGTAGGAGAT GCCTACAATC TTCCCAACAA AGACGAGTTA CTTACACTCA TTCGAGGAAA GAAGATCTTC 4640

TCTTCCTTCG ACTGTAAGTC AGGATTCTGG CAAGTTCTGC TAGATCAAGA ATCAAGACCT CTAACGGCAT TCACATGTCC 4720

ACAAGGTCAC TACGAATGGA ATGTGGTCCC TTTCGGCTTA AAGCAAGCTC CATCCATATT CCAAAGACAC ATGGACGAAG 4800

CATTTCGTGT GTTCAGAAAG TTCTGTTGCG TTTATGTCGA CGACATTCTC GTATTCAGTA ACAACGAAGA AGATCATCTA 4880

CTTCACGTAG CAATGATCTT ACAAAAGTGT AATCAACATG GAATTATCCT TTCCAAGAAG AAAGCACAAC TCTTCAAGAA 4960

GAAGATAAAC TTCCTTGGTC TAGAAATAGA TGAAGGAACA CATAAGCCTC AAGGACATAT CTTGGAACAC ATCAACAAGT 5040

TCCCCGATAC CCTTGAAGAC AAGAAGCAAC TTCAGAGATT CTTAGGCATA CTAACATATG CCTCGGATTA CATCCCGAAG 5120

CTAGCTCAAA TCAGAAAGCC TCTGCAAGCC AAGCTTAAAG AAAACGTTCC ATGGAGATGG ACAAAAGAGG ATACCCTCTA 5200

CATGCAAAAG GTGAAGAAAA ATCTGCAAGG ATTTCCTCCA CTACATCATC CCTTACCAGA GGAGAAGCTG ATCATCGAGA 5280

CCGATGCATC AGACGACTAC TGGGGAGGTA TGTTAAAAGC TATCAAAATT AACGAAGGTA CTAATACTGA GTTAATTTGC 5360

AGATACGCAT CTGGAAGCTT TAAAGCTGCA GAAAAGAATT ACCACAGCAA TGACAAAGAG ACATTGGCGG TAATAAATAC 5440

TATAAAGAAA TTTAGTATTT ATCTAACTCC TGTTCATTTT CTGATTAGGA CAGATAATAC TCATTTCAAG AGTTTCGTTA 5520

ATCTCAATTA CAAAGGAGAT TCGAAACTTG GAAGAAACAT CAGATGGCAA GCATGGCTTA GCCACTATTC ATTTGATGTT 5600

GAACACATTA AAGGAACCGA CAACCACTTT GCGGACTTCC TTTCAAGAGA ATTCAATAAG GTTAATTCCT AATTGAAATC 5680

CGAAGATAAG ATTCCCACAC ACTTGTGGCT GATATCAAAA GGCTACTGCC TATTTAAACA CATCTCTGGA GACTGAGAAA 5760

ATCAGACCTC CAAGCATGGA GAACATAGAA AAACTCCTCA TGCAAGAGAA AATACTAATG CTAGAGCTCG ATCTAGTAAG 5840

AGCAAAAATA AGCTTAGCAA GAGCTAACGG CTCTTCGCAA CAAGGAGACC TCTCTCTCCA CCGTGAAACA CCGGAAAAAG 5920

AAGAAGCAGT TCATTCTGCA CTGGCTACTT TTACGCCATC TCAAGTAAAA GCTATTCCAG AGCAAACGGC TCCTGGTAAA 6000

GAATCAACAA ATCCGTTGAT GGCTAATATC TTGCCAAAAG ATATGAATTC AGTTCAGACT GAAATTAGGC CCGTAAAGCC 6080

ATCGGACTTC TTACGTCCAC ATCAGGGAAT TCCAATCCCA CCAAAACCTG AACCTAGCAG TTCAGTTGCT CCTCTCAGAG 6160

ACGAATCGGG TATTCAACAC CCTCATACCA ACTACTACGT CGTGTATAAC GGACCTCATG CCGGTATATA CGATGACTGG 6240

GGTTGTACAA AGGCAGCAAC AAACGGTGTT CCCGGAGTTG CGCATAAGAA GTTTGCCACT ATTACAGAGG CAAGAGCAGC 6320

AGCTGACGCG TATACAACAA GTCAGCAAAC AGATAGGTTG AACTTCATCC CCAAAGGAGA AGCTCAACTC AAGCCCAAGA 6400

GCTTTGCGAA GGCCTTAACA AGCCCACCAA AGCAAAAAGC CCACTGGCTC ATGCTAGGAA CTAAAAAGCC CAGCAGTGAT 6480

CCAGCCCCAA AAGAGATCTC CTTTGCCCCA GAGATCACAA TGGACGACTT CCTCTATCTC TACGATCTAG TCAGGAAGTT 6560

CGACGGAGAA GGTGACGATA CCATGTTCAC CACTGATAAT GAGAAGATTA GCCTTTTCAA TTTCAGAAAG AATGCTAACC 6640

CACAGATGGT TAGAGAGGCT TACGCAGCAG GTCTCATCAA GACGATCTAC CCGAGCAATA ATCTCCAGGA GATCAAATAC 6720

CTTCCCAAGA AGGTTAAAGA TGCAGTCAAA AGATTCAGGA CTAACTGCAT CAAGAACACA GAGAAAGATA TATTTCTCAA 6800

GATCAGAAGT ACTATTCCAG TATGGACGAT TCAAGGCTTG CTTCACAAAC CAAGGCAAGT AATAGAGATT GGAGTCTCTA 6880

AAAAGGTAGT TCCCACTGAA TCAAAGGCCA TGGAGTCAAA GATTCAAATA GAGGACCTAA CAGAACTCGC CGTAAAGACT 6960

GGCGAACAGT TCATACAGAG TCTCTTACGA CTCAATGACA AGAAGAAAAT CTTCGTCAAC ATGGTGGAGC ACGACACGCT 7040

```
TGTCTACTCC AAAAATATCA AAGATACAGT CTCAGAAGAC CAAAGGGCAA TTGAGACTTT TCAACAAAGG GTAATATCCG   7120
GAAACCTCCT CGGATTCCAT TGCCCAGCTA TCTGTCACTT TATTGTGAAG ATAGTGGAAA AGGAAGGTGG CTCCTACAAA   7200
TGCCATCATT GCGATAAAGG AAAGGCCATC GTTGAAGATG CCTCTGCCGA CAGTGGTCCC AAAGATGGAC CCCCACCCAC   7280
GAGGAGCATC GTGGAAAAAG AAGACGTTCC AACCACGTCT TCAAAGCAAG TGGATTGATG TGATATCTCC ACTGACGTAA   7360
GGGATGACGC ACAATCCCAC TATCCTTCGC AAGACCCTTC CTCTATATAA GGAAGTTCAT TTCATTTGGA GAGGACACGC   7440
TGAAATCACC AGTCTCTCTC TACAAATCTA TCTCTCTCTA TAATAATGTG TGAGTAGTTC CCAGATAAGG GAATTAGGGT   7520

TCTTATAGGG TTTCGCTCAT GTGTTGAGCA TATAAGAAAC CCTTAGTATG TATTTGTATT TGTAAAATAC TTCTATCAAT   7600
AAAATTTCTA ATTCCTAAAA CCAAAATCCA GTACTAAAAT CCAGATCTCC TAAAGTCCCT ATAGATCTTT GTGGTGAATA   7680
TAAACCAGAC ACGAGACGAC TAAACCTGGA GCCCAGACGC CGTTTGAAGC TAGAAGTACC GCTTAGGCAG GAGGCCGTTA   7760
GGGAAAAGAT GCTAAGGCAG GGTTGGTTAC GTTGACTCCC CCGTAGGTTT GGTTTAAATA TCATGAAGTG GACGGAAGGA   7840
AGGAGGAAGA CAAGGAAGGA TAAGGTTGCA GGCCCTGTGC AAGGTAAGAC GATGGAAATT TGATAGAGGT ACGTTACTAT   7920
ACTTATACTA TACGCTAAGG GAATGCTTGT ATTTACCCTA TATACCCTAA TGACCCCTTA TCGATTTAAA GAAATAATCC   8000
                      Gap 1
GCATAAGCCC CCGCTTAAAA AATT                                                                8024
```

Figure 1. Nucleotide Sequence of Circular CaMV DNA (Isolate Cabb B-S)

Only the sequence of the β strand, which has two gaps and the same polarity as viral mRNA, is presented. Numbering begins with the first dG at gap 1. The sequence is written as a continuous chain everywhere complementary to the transcribed α strand in order to facilitate consideration of coding capacity but, in fact, the β strand is broken at gaps 2 and 3 and the extremities of the sequence at each gap are redundant (see text and Figure 6). The asterisk marks a dG residue which may be paired with a modified dC residue in the complementary strand.

Since the CaMV isolate used to establish the sequence has not been cloned, we were prepared to find some sequence variation within the viral DNA population. Rather surprisingly, however, not one clear-cut example of such heterogeneity was detected in the two CaMV DNA preparations used to establish the sequence. Very occasionally, double signals occurred at a given position in a sequence gel, but examination of other gels covering the region in question or its complementary strand generally revealed such signals to be spurious. Thus, although the existence of rare sequence variants affecting a small proportion of the DNA population cannot be ruled out, most of the DNA molecules of Cabb B-S isolate conform to the sequence presented in Figure 1. Restriction analysis of a large number of pBR322-CaMV DNA recombinants has revealed little sequence variation among the clones, a result consistent with this conclusion (Hohn et al., 1980).

The Coding Region

While it is evident that the nucleotide sequence will ultimately tell us much if not all about the properties of the CaMV genome, a straightforward extrapolation from the DNA sequence to the properties of the final gene products, the viral proteins, is rendered difficult by the complexity of mRNA maturation in eucaryotes and our lack of knowledge of the signals governing this process. Nevertheless, analysis of the coding capacity of the nucleotide sequence permits the broad outlines of CaMV genetic organization, if not the details, to be discerned.

Transcription of CaMV DNA is asymmetric. There is agreement that virus-specified RNA present in infected turnip leaves (Hull et al., 1979; Al Ani et al., 1980) or protoplasts (Howell and Hull, 1978) hybridizes exclusively with the α strand of CaMV DNA, that is, the strand containing only one gap. It is reasonable to believe that most or all of these RNA transcripts encode viral proteins. Such a view is in fact borne out by examination of th coding capacity of the nucleotide sequenc . Figure 2a presents an analysis of th distribution of TGA, TAG and TAA t rmination codons in th three possible coding frames of th β strand, the sequ nce having the same polarity as would an

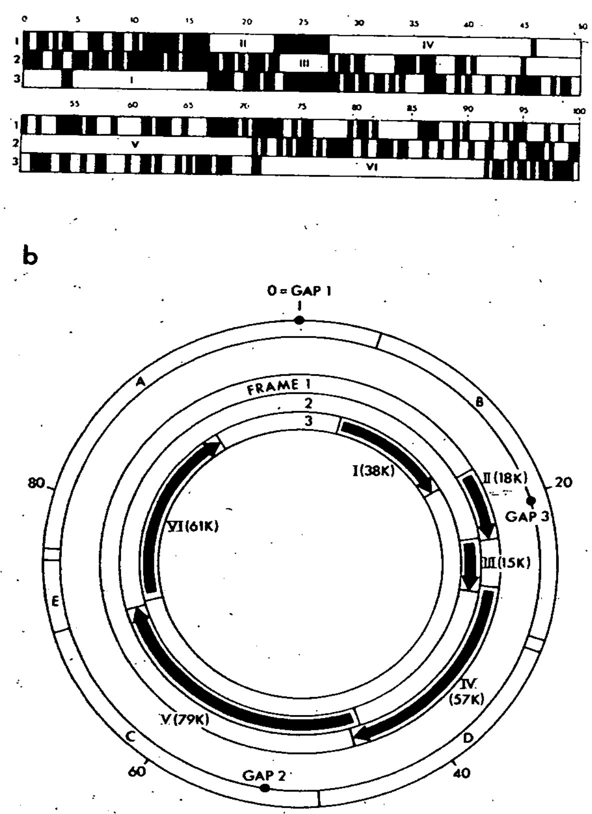


Figure 2a. Distribution of Nonsense Codons in the CaMV β Strand Sequence

Triplet frame 1 begins with the first residue in the sequence as presented in Figure 1, frame 2 with the second nucleotide and frame 3 with the third residue. The sequence in each coding frame was divided into consecutive segments of 12 triplet codons (36 nucleotides). When a nonsense codon occurs in a segment, the corresponding rectangle in the diagram is blackened (▇).

Figure 2b. Distribution of Potential Coding Regions on the Circular CaMV DNA Map

Inner circles give the positions of the six long open reading frames identified in (a) along with the molecular weight in kilodaltons of the longest possible translation product for which each could code (assuming that translation begins with first in-phase ATG initiation codon in each open region). The outer circle gives the positions of Eco RI fragments A–E and the three gaps.

RNA transcript of the α strand. The diagram reveals that, apart from a region of about 1000 nucleotides in the vicinity of gap 1 (91–4 map units), virtually the entire sequence is free of nonsense codons for considerable distances in one or another of the three reading frames. The longest such potential coding sequence, region V, is 2082 nucleotides long. Regions IV and VI are each more than 1500 residues in length, while region I is 1000 nucleotides, region II 500 nucleotides and region III 400 nucleotides long (Table 1). By way of contrast, the sequence complementary to the β strand, a sequence which is not transcribed and hence almost certainly does not code, contains no uninterrupted triplet reading frame of more than 370 nucleotides. It is noteworthy that, with the exception of about 120 nucleotides at the junction between regions IV and V, there is little overlap between successive coding regions. Thus CaMV has not economized its genetic information as has ϕX174 (Sanger et al., 1977) by using different reading frames of the same sequence to specify different proteins.

We are confident that the six long uninterrupted reading frames identified in Figure 2a represent the effective protein coding potential of CaMV DNA. The maximum lengths of the polypeptides for which the six regions may code (assuming translation starts with the first in-phase AUG in each region and that there is no read-through of termination codons) are shown in Figure 2b. Whether any or all of these polypeptides are in fact synthesized in infected tissue is uncertain and will probably remain so until more is known about the manner in which viral RNA transcripts are processed. Evidence is presented below, however, suggesting that a part of region IV encodes the major viral coat protein.

Virus-specific RNA transcripts synthesized in protoplasts prepared from CaMV-infected turnip leaves have been reported to be quite large, with molecular weights in the range of 2×10^6 daltons or greater (Howell and Hull, 1978; Howell, Odell and Dudley, 1979). RNA molecules of this length could accommodate essentially all of the coding portion of the CaMV genome (regions I–VI). Transcripts isolated directly from turnip leaves, on the other hand, were found to be significantly smaller, sedimenting as distinct 18S and 25S species in sucrose gradients (Howell et al., 1979). Howell et al. (1979) have suggested that protoplasts are defective in viral mRNA processing and that the large RNA transcripts which accumulate in this system may be precursors of the smaller molecules. It is uncertain, however, whether this maturation process would involve the joining together (splicing) of discontinuous regions of the mRNA precursor as observed for many eucaryotic mRNAs. Howell et al. (1979) argue that splicing probably occurs, based upon their analysis of hybridization patterns in Southern blots between the radioactive 18S and 25S viral mRNAs and CaMV restriction fragments. In our laboratory, however, electron microscopic examination of specific RNA-DNA heteroduplexes between CaMV DNA and polyadenylated RNA fractions isolated from infected turnip leaves has so far failed to reveal a single instance of single-stranded DNA looping out of such hybrids, as would be expected if the DNA and RNA molecules are not colinear (J. Menissier, personal communication).

Examination of the nucleotide sequence itself provides two, admittedly weak, arguments against extensive involvement of splicing in the mRNA maturation process. First, the canonical splice-junction sequences (5′) $^{A}_{C}$AG↓GTAAGT ... TYTYYYTXCAG↓G (Lerner et al., 1980; Y is a pyrimidine, X is any nucleotide) or close variants thereof are rarely found in the CaMV β strand sequence and then not in places where a splice might be expected, such as across regions where there is a shift in the coding frame. It may be objected, however, that the splice junctions of plants and their viruses may be sufficiently different from those of animals and insects to escape recognition.

A second argument rests upon the way in which the reading frame in the coding region jumps abruptly from one phase to another so that, in all but one case (the junction between regions V and VI), successive open regions overlap slightly or are separated from

Table 1. Coordinates of Possible Coding Regions of CaMV DNA

Open Region	Start: Nucleotide	Start: Map Unit	End: Nucleotide	End: Map Unit	First ATG Nucleotide	Protein Molecular Weight (Kilodaltons)[a]
I	331	4.12	1347	16.79	364	38
II	1328	16.55	1828	22.78	1349	18
III	1812	22.58	2219	27.65	1830	15
IV	2168	27.02	3670	45.74	2201	57
V	3591	44.75	5672	70.69	3633	79
VI	5713	71.20	7338	91.45	5776	61

[a] Assuming polypeptide chain synthesis starts with the first in-phase AUG in each open region.

one another by only a few nucleotides (Table 1). Thus it is unlikely that the CaMV genome contains noncoding intervening sequences (introns) in the primary coding region (map units 4–91) which are present in primary RNA transcripts but eliminated from mature mRNAs. A splicing pattern of the type observed for adenovirus late mRNAs, however, in which a single nontranslated 5′ leader sequence is spliced to each of several alternate coding sequences (Philipson, 1979), cannot be ruled out from the sequence data alone.

Viral Gene Products

Additional insight into the organization of the CaMV genome could be gained if we could unequivocally equate one or more of the potential coding regions described in the preceding section with a viral gene product. One obvious candidate for consideration is the major viral structural protein. Early investigations of the architecture of CaMV virions were rather confused, with estimations of the number of structural polypeptides present varying from two to seven or more (Tezuka and Taniguchi, 1972; Kelly, Cooper and Walkey, 1974; Brunt et al., 1975; Hull and Shepherd, 1976). The molecular weights of the various polypeptides ranged from 30 to 85 kilodaltons while their molar ratios often depended upon the method of virus purification and the age of the virus preparation. Al Ani, Pfeiffer and Lebeurier (1979), however, have recently shown that the situation is in reality much simpler. They argue that there is only one major CaMV structural protein, a polypeptide of about 42 kilodaltons that we shall refer to as P42. Smaller polypeptides normally associated with virion preparations were shown to have sequences in common with P42 and no doubt arise by proteolytic degradation of the major species, while the polypeptides in the 70–80 kilodalton range are probably artifactual dimers of P42 and its degradation products arising from incomplete reduction of disulfide bonds (Al Ani et al., 1979).

A distinctive feature of CaMV coat protein is its unusually high lysine composition, which amounts to 18% (on a molar basis) of the amino acid content of total virion protein (Brunt et al., 1975). We have examined the coding capacity of each of the six putative coding regions identified in Figure 2 and find that only one, region IV, has the potential to code for a protein approaching this degree of richness in lysine. Figure 3a presents the amino acid sequence corresponding to region IV, beginning with the first ATG initiation codon in the appropriate phase. This hypothetical polypeptide can be seen to have an extremely lysine-rich region n ar the carboxy t rminus (amino acid r sidues 333–410). Table 2 gives the amino acid composition calculat d for a polypeptide spanning this lysine-rich core and having a molecular weight of about 42 kilodaltons. The xact boundaries of the putativ coat protein polyp ptide were chosen to optimize the fit, which is excellent, between the calculated values and those observed for viral coat protein (Brunt et al., 1975), but may be adjusted slightly without affecting the figures significantly. The fit is all the more striking if one takes into account the fact that the amino acid analysis was performed on total virion protein, which no doubt included degradation products of the basic 42 kilodalton polypeptide.

Efforts to elicit synthesis of P42 in cell-free systems primed with mRNA fractions from CaMV-infected leaves have not been successful. Such mRNA fractions do, however, direct synthesis of a virus-specific polypeptide with an estimated molecular weight of about 55 kilodaltons (P55) in a rabbit reticulocyte cell-free system (R. Al Ani and A. Lesot, personal communication). We consider it probable that P55 is the product of total translation of region IV, which could give rise to a polypeptide in this size range (Figure 2b). Viral coat protein, then, would be synthesized in the form of a precursor, thus explaining our failure to obtain mature coat protein in in vitro translation systems. Experiments are under way to discover if there is in fact a serological relationship between in vitro synthesized P55 and viral coat protein.

The lysine-rich core of P42 presumably interacts with the DNA in the intact virion. In this regard, it is noteworthy that the extremities of the longer sequence are extremely rich in glutamic and aspartic acid residues (Figure 3). If the precursor-product relation put forward above for P55 and P42 proves correct, then the supplementary acidic residues in the precursor polypeptide may serve to neutralize the lysine-rich core in the absence of DNA. Processing of the coat protein precursor would, by eliminating the acidic terminal sequences of the longer polypeptide, leave the lysine-rich core of the mature coat protein free to interact with DNA.

The only other CaMV gene product that has been unambiguously identified is a polypeptide of about 62 kilodaltons (P62) which is the most prominent virus-specified translation product primed by polyadenylated mRNA from CaMV-infected leaves (Al Ani et al., 1980). An apparently similar if not identical translation product with an estimated molecular weight of 66 kilodaltons has been described by Howell et al. (1979). P62 can be detected in protein extracted from infected leaves, and cell fractionation experiments indicate that it is associated with the inclusion bodies known as viroplasts (Howell et al., 1979; Al Ani et al., 1980), which accumulate in the cytoplasm of infected cells (for references see Shepherd, 1979). Howell et al. (1979) have used the hybrid-arrested translation (HART) technique (Paterson, Roberts and Kuff, 1977) and cloned Eco RI fragments to show that the greatest part of the sequence encoding their 66 kilodalton polypeptide lies within Eco RI fragment A (map positions 76–5; Figure 2b). Thus open region VI, which is mostly contained within fragment A and has the po-

a

1 5 10 15 20 25 30

fmet Ala Glu Ser Ile Leu Asp Arg Thr Ile Asn Arg Phe Trp Tyr Asn Leu Gly Glu Asp Cys Leu Ser Glu Ser Gln Phe Asp Leu Met

Ile Arg Leu Met Glu Glu Ser Leu Asp Gly Asp Gln Ile Ile Asp Leu Thr Ser Leu Pro Ser Asp Asn Leu Gln Val Glu Gln Val Met

Thr Thr Thr Glu Asp Ser Ile Ser Glu Glu Glu Ser Glu Phe Leu Leu Ala Ile Gly Glu Thr Ser Glu Glu Glu Ser Asp Ser Gly Glu

Glu Pro Glu ▼ Phe Glu Gln Val Arg Met Asp Arg Thr Gly Gly Thr Glu Ile Pro Lys Glu Glu Asp Gly Glu Gly Pro Ser Arg Tyr Asn

Glu Arg Lys Arg Lys Thr Pro Glu Asp Arg Tyr Phe Pro Thr Gln Pro Lys Thr Ile Pro Gly Gln Lys Gln Thr Ser Met Gly Met Leu

Asn Ile Asp Cys Gln Thr Asn Arg Arg Thr Leu Ile Asp Asp Trp Ala Ala Glu Ile Gly Leu Ile Val Lys Thr Asn Arg Glu Asp Tyr

Leu Asp Pro Glu Thr Ile Leu Leu Leu Met Glu His Lys Thr Ser Gly Ile Ala Lys Glu Leu Ile Arg Asn Thr Arg Trp Asn Arg Thr

Thr Gly Asp Ile Ile Glu Gln Val Ile Asp Ala Met Tyr Thr Met Phe Leu Gly Leu Asn Tyr Ser Asp Asn Lys Val Ala Glu Lys Ile

Asp Glu Gln Glu Lys Ala Lys Ile Arg Met Thr Lys Leu Gln Leu Cys Asp Ile Cys Tyr Leu Glu Glu Phe Thr Cys Asp Tyr Glu Lys

Asn Met Tyr Lys Thr Glu Leu Ala Asp Phe Pro Gly Tyr Ile Asn Gln Tyr Leu Ser Lys Ile Pro Ile Ile Gly Glu Lys Ala Leu Thr

Arg Phe Arg His Glu Ala Asn Gly Thr Ser Ile Tyr Ser Leu Gly Phe Ala Ala Lys Ile Val Lys Glu Glu Leu Ser Lys Ile Cys Asp

Leu Ser Lys Lys Gln Lys Lys Leu Lys Lys Phe Asn Lys Lys Cys Cys Ser Ile Gly Glu Ala Ser Thr Glu Tyr Gly Cys Lys Lys Thr

Ser Thr Lys Lys Tyr His Lys Lys Arg Tyr Lys Lys Lys Tyr Lys Ala Tyr Lys Pro Tyr Lys Lys Lys Lys Lys Phe Arg Ser Gly Lys

Tyr Phe Lys Pro Lys Glu Lys Lys Gly Ser Lys Gln Lys Tyr Cys Pro Lys Gly Lys Lys Asp Cys Arg Cys Trp Ile Cys Asn Ile Glu

Gly His Tyr Ala Asn Glu Cys Pro Asn Arg Gln Ser Ser Glu Lys Ala His Ile Leu Gln Gln Ala Glu Lys Leu Gly Leu Gln ▼ Pro Ile

Glu Glu Pro Tyr Glu Gly Val Gln Glu Val Phe Ile Leu Glu Tyr Lys Glu Glu Glu Glu Glu Thr Ser Thr Glu Glu Ser Asp Gly Ser

Ser Thr Ser Glu Asp Ser Asp Ser Asp Umber

b

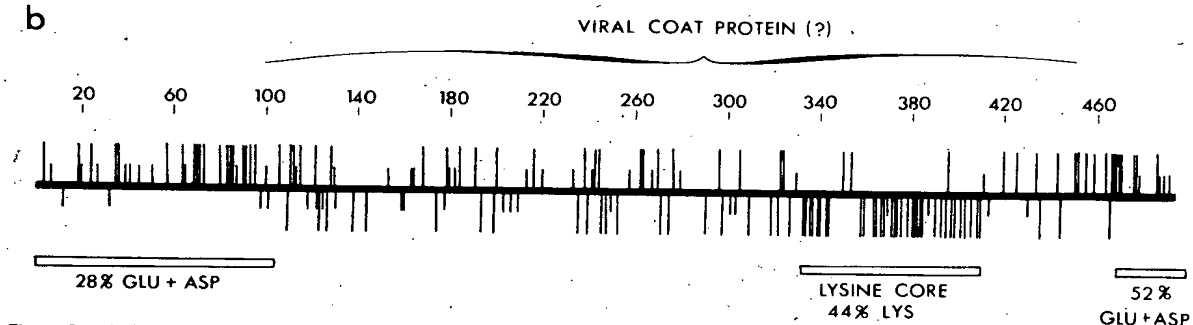


Figure 3a. Amino Acid Sequence Coded for by Open Region IV

The sequence begins with the first in-phase ATG codon (2201–2203) and proceeds to the end of the open region. The two inverted triangles mark the beginning and the end of the portion of the sequence which may correspond to viral coat protein.

Figure 3b. Distribution of Acidic and Basic Amino Acid Residues in the Open Region IV Amino Acid Sequence

Half- and full-length vertical lines pointing upward denote aspartic acid and glutamic acid residues, respectively, while half- and full-length lines pointing downward denote asparagine and lysine.

tential to code for a protein of about the right size, would seem to be the best candidate for the viroplast protein cistron. Al Ani et al. (1980) observed that hybridization of mRNA fractions with both Eco RI fragments A and E inhibits in vitro synthesis of P62, consistent with the aforesaid localization (Figure 2b). The same investigators also found, however, that hybridization with Eco RI fragment B (map units 5–30) also inhibited P62 synthesis, a result not easily reconciled with the sequence data unless splicing is invoked. Reinvestigation of this matter using cloned Eco RI fragments to eliminate the possibility of contamination of fragment B by fragment A appears essential.

The Noncoding Region

The sequence separating the end of coding region VI from the beginning of coding region I does not appear to encode protein. Reading frames 1 and 2 in this portion of the sequence (map units 91–4) are blocked by numerous termination codons (Figure 2). Frame 3 contains a small open region of about 300 nucleotides (map units 0–4) but there is no in-phase ATG initiation codon. The data of Hull et al. (1979) suggest that transcription commences within a few map units downstream of gap 1, the zero point in our sequence, and terminates somewhere between map units 76 and 100. This scheme fits in well with the sequence, as we would expect transcription to begin at or before map position 4.1, the beginning of coding region I (Table 1), and proceed to at least map position 91, the end of coding region VI. About 250 nucleotides downstream from the final triplet in region VI appears the sequence AATAAA (7598–7603), whose RNA equivalent is to be found 15–30 nucleotides prior to the poly(A) tail in a great many eucaryotic mRNAs (Proudfoot and Brownlee, 1976). If this feature is similarly located in CaMV RNA transcripts, then the

Table 2. Amino Acid Composition of Cauliflower Mosaic Virus Coat Protein

		Molar %	
Amino Acid	N° Residues[a]	Calculated	Observed[b]
Lys	56	15.91	17.97
His	5	1.42	1.04
Arg	19	5.40	4.84
Asp + Asn	32	9.09	8.99
Thr	23	6.53	6.57
Ser	16	4.54	4.72
Glu + Gln	46	13.07	11.52
Pro	13	3.69	3.46
Gly	21	5.96	6.91
Ala	16	4.54	5.07
Cys	13	3.69	2.88
Val	5	1.42	1.38
Met	8	2.27	1.61
Ile	26	7.39	6.80
Leu	23	6.53	7.49
Tyr	20	5.68	5.65
Phe	10	2.84	3.11
Trp	3	ND[c]	ND[c]
Total	355 = 41,417 daltons		

[a] Taken from Figure 3.
[b] From Brunt et al., 1975.
[c] Tryptophan content was not measured.

termination point for transcription would fall near map position 95.

With regard to initiation of RNA transcription, many eucaryotic mRNA coding genes have AT-rich regions (typically, a variant of the sequence TATAAAA), often flanked by GC-rich sequences, 20–30 nucleotides upstream of the transcription initiation point (for references see Benoist et al., 1980). The CaMV DNA sequence between gap 1 and the beginning of coding region I contains several AT-rich regions but perhaps th most eye-catching example occurs just *before* gap 1: GCCCCCGCTTAAAAAATT (residues 8007–8024). For the present, we reserve judgment on the role, if any, of such sequences in transcription initiation until the 5′ terminus of the initial RNA transcript has been characterized.

Th Gaps

Determination of the sequence in the vicinity of the gaps presented special problems. We have observed, as have Volovitch et al. (1979), that restriction fragments containing a gap often display perturbed behavior during electrophoresis, migrating as a diffuse band or family of bands in the gel. We have also observed that proximity to a gap may render certain restriction enzyme sites refractory to attack. Hohn et al. (1980) have shown that CaMV DNA cloned in pBR322 (no gaps) possesses a Hind III restriction site within 110 nucleotides of gap 3 that was not detected in noncloned DNA. Sequence analysis has shown that the Hind III recognition sequence in fact exists in noncloned viral DNA (positions 1513–1518), but only once were we successful in obtaining cleavage at this site. Several examples of incomplete cleavage at Taq I sites in the vicinity of gaps were also noted.

In spite of these difficulties, a number of short 5′ ^{32}P-labeled double-stranded restriction fragments encompassing each of the three gaps were isolated. When such a fragment was subjected to strand separation, three radioactive single-stranded fragments of disparate length were produced. The two smaller fragments correspond to the two segments of the strand interrupted by the gap while the longest fragment is the continuous complementary strand. Evidently, one of the two shortened single-stranded fragments will have the 5′ terminal nucleotide of the gap at its labeled extremity so that sequence data from this fragment should exactly situate the 5′ limit of the gap with respect to the sequence of the complementary strand. The sequence of the other shortened fragment, which has the restriction site as 5′ terminus and the gap at its 3′ extremity, should in principle define the approximate 3′ limit of the gap, providing that the restriction site is sufficiently close to the gap so that the sequence can be read to its end.

5′ Extremities of the Gaps

Sequence gels for 5′ labeled fragments originating from gaps 1, 2 and 3 are shown in Figure 4. Hull et al. (1979) have identified the 5′ terminal nucleotides of the gaps in CaMV DNA (isolate Cabb B-JI) as dA for gaps 1 and 2 and dG for gap 3. We find the same 5′ termini for the gaps in the DNA of our isolate (Cabb B-S) both upon sequencing gels (Figure 4) or by total P1 nuclease digestion of the 5′ ^{32}P-labeled fragments and electrophoresis of the digestion products at pH 3.5 (data not shown). In numerous experiments, however, the signal corresponding to the second nucleotide from the 5′ labeled end of each gap was always obscured by a heavily labeled diffuse band or pair of bands traversing all four lanes of the sequencing gel (Figure 4). Anomalous signals of this sort were never encountered for ordinary restriction fragments prepared and electrophoresed in parallel, suggesting that an unusual structure or modified nucleotide may be present at these positions in the DNA molecule. The nature of these unusual residues is currently under investigation.

In the course of sequence determination we observed that gap 2 is absent from a portion of the DNA molecules of our preparations. A short Taq I restriction fragment spanning the region of gap 2 was found to migrate as two distinct approximately equimolar components in polyacrylamide gels. Characterization of

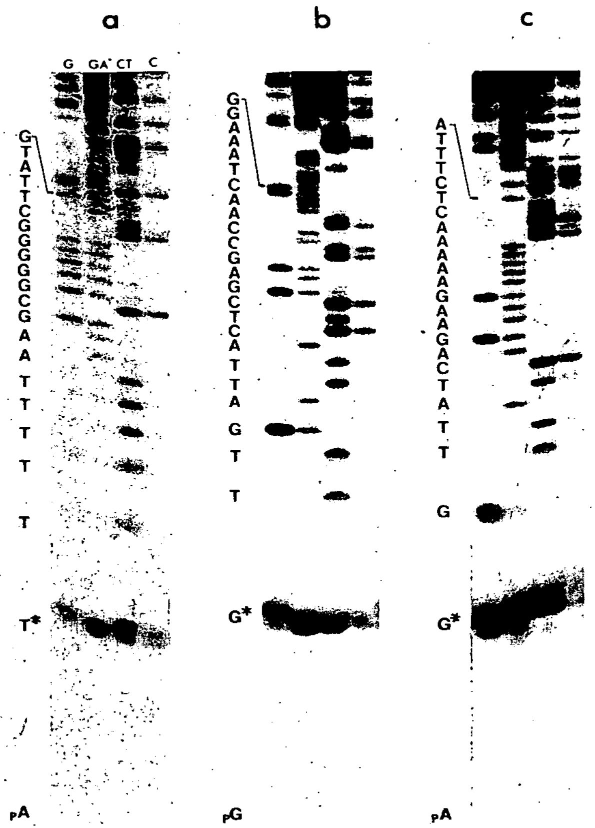


Figure 4. Maxam-Gilbert 20% Sequence Gels of Single-Stranded Eco RI Fragments 5′ ^{32}P-Labeled at the Gaps

The signal for the first nucleotide in the sequence is not readily visible in the reproduction. Asterisk indicates nucleotides which may be unusual or modified. (a) Gap 1; (b) gap 3; (c) gap 2.

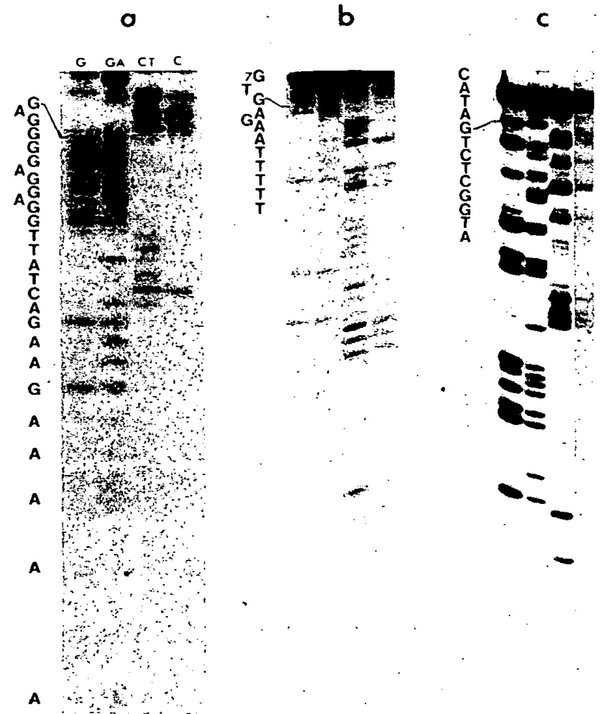


Figure 5. Maxam-Gilbert Sequence Gels Showing the 3′ Termini of the Gaps

(a) Single-stranded Eco RI fragment 3′ ^{32}P-labeled at gap 2; position of first signal corresponds to a dinucleotide (20% gel); (b) single-stranded 5′ ^{32}P-labeled Hinf I fragment with 3′ terminus at gap 3 (8% gel); (c) single-stranded 5′ ^{32}P-labeled Bgl II fragment with 3′ terminus at gap 1 (8% gel).

these fragments revealed that the more slowly migrating of the two species contained gap 2, giving rise to three 5′ labeled single-stranded fragments upon strand separation, two of which originated from the restriction cuts and the third from the gap, whereas the other fragment, although otherwise identical in sequence, consisted only of the two uninterrupted complementary strands.

3′ Extremities of the Gaps

Volovitch et al. (1979) have reported that homopolymer tracts may be added at all three gaps in native CaMV DNA with terminal deoxynucleotidyl transferase, indicating that the 3′ terminal nucleotide of each discontinuity has a free 3′ OH group, but only in the case of gap 2 have we succeeded in introducing enough label at the 3′ terminal position for sequencing purposes. Nevertheless, a fairly precise localization of the 3′ limits of gaps 1 and 3 was obtained by sequencing short 5′ labeled restriction fragments having the gaps at their 3′ termini. (An appropriat fragment t rminating in gap 2 exists but could not be isolat d in quantities sufficient for sequence analysis.) Starting from a Bgl II site about one hundred residues upstream from gap 1 on the 5′ side (with respect to the discontinuous strand), we were able to read the sequence toward gap 1 for all but the last one or two residues at the 3′ end of this fragment, where the heavily labeled band of undegraded material obscured the specific signals in the sequence ladder (Figure 5c). The sequence so determined extends to within two residues of the 5′ terminal nucleotide of the gap (Figure 6). Thus the single-stranded region separating the 5′ and 3′ ends of the gap 1 discontinuity is no more than one or two nucleotides in length.

Figure 5b shows a sequence ladder for a 5′ labeled Hinf I fragment terminating at gap 3. The fragment has the sequence TTTTTAAGAGTGGGGGGG. . . at its 3′ extremity (the seven dG signals in the final run are not readily visible in the reproduction but can be seen on the original film). Comparing this to the sequence of the complementary strand reveals that, surprisingly enough, the 3′ terminal sequence at the discontinuity *overlaps* the first two residues of the 5′ terminal continuation of the strand (Figure 6).

An even more sizable sequence overlap exists at gap 2. As mentioned above, we were successful in

```
                 8010                   1            10              20
GAP 1    5'-AGCCCCCGCTTAAAAAATTGGTATCAGAGCCATGAATCGG-
           -TCGGGGGCGAATTTTTTA  CATAGTCTCGGTACTTAGCC-5'
                            *

                                    *
                            5'-AGGTTATCAGAAGAAAAA
                                                  C
GAP 2    5'-TATTCTTTCAGAGGGGAGGAGGTTATCAGAAGAAAAA    TC-
                                                 OH
           -ATAAGAAAGTCTCCCCTCCTCCAATAGTCTTCTTTTTGAG-5'
            4200                            4230

                            5'-GG*
                                  T
GAP 3    5'-CCATTTTTAAGAGTGGGGGGG  TGATTACTCGAGCCAACT-
           -GGTAAAAATTCTCACCCCCCCAACTAATGAGCTCGGTTGA-5'
                 1620             1640
```

Figure 6. Sequence in the Vicinity of the Gaps

The sequence corresponding to the β strand is written on the upper line and the complementary α strand on the lower line. Asterisk denotes unusual or modified nucleotide (see text). Dashed arrows indicate the extent to which sequence could be read starting from 5′ labeled restriction cuts upstream of the gap.

fixing a ^{32}P–AMP residue to the 3′ terminus of this gap by incubating Eco RI fragment C with α–^{32}P–ATP and terminal transferase. The fragment with the gap at its 3′ terminus was purified by strand separation and sequenced (Figure 5a). When read in a 5′ to 3′ sense, the last 19 residues of the 3′ terminal sequence of this fragment were found to be identical to the first 19 residues of the 5′ continuation of the strand (compare Figure 5a to Figure 4c). As there is no corresponding sequence duplication in the complementary strand, it follows that one or the other of the redundant sequences must branch off of the double helix as a single-stranded tail, as shown in Figure 6. We suggest that the displaced strand is normally that possessing the free 5′ OH extremity because of the relative ease with which the 5′ extremities of the discontinuities can be labeled with polynucleotide kinase (our unpublished observations).

Figure 6 summarizes the sequence in the vicinity of the three gaps. It can be seen that, while none of the sequences are identical, the 5′ termini of gaps 2 and 3 both fall in regions in which the complementary strand is very rich in C: CCCCCCC (1634–1628) for gap 3 and CCTCCTCCCC (4220–4211) for gap 2. These are the two pyrimidine tracts having the highest C content in the entire molecule. Gap 1 is also close to a C-rich sequence, CCCCCGC (8008–8014), but is separated from it by the symmetric AT tract TTAAAAAATT (8015–8024) mentioned above. The differences in sequence around the gaps presumably reflect differences in function. If so, the proximity of gap 1 to the beginning of the coding region leads naturally to the idea that it may be involved in initiation of RNA transcription, but plausible roles for gaps 2 and 3 come less easily to mind. One interesting possibility is that these gaps are start/stop points for replication of viral DNA, as it is evident that redundant terminal sequences like those associated with gaps 2 and 3 could arise if a round of DNA replication proceeds for a short way beyond the original starting point.

Experimental Procedures

Details of the procedure for propagation of CaMV (isolate Cabb B-S), purification of the virus by the Triton X-100–urea method (Hull, Shepherd and Harvey, 1976) and extraction of the DNA have been given elsewhere (Hohn et al., 1980). Digestion of the DNA with restriction enzymes was performed as recommended by the supplier (Biolabs). After restriction, the fragments were dephosphorylated by incubation for 2 hr at 50°C with 0.1–0.5 units of E. coli alkaline phosphatase (Boehringer) per μg DNA. The reaction was stopped by emulsifying with phenol, phenol was eliminated by extraction with ether and the fragments were precipitated with ethanol. The DNA fragments were 5′ end-labeled by incubation at 37°C for 30 min with 0.5 units polynucleotide kinase (Boehringer) per μg DNA in the presence of a 2–5 fold molar excess (with respect to DNA 5′ termini) of γ–^{32}P–ATP (3000 Ci/mmole; Amersham). The reaction was carried out in 70 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM spermidine, 0.1 mM EDTA, 5 mM dithiothreitol and 25% glycerol. If fragments in which the 5′ termini were flush or recessed were to be labeled, 25% dimethylsulfoxide was included in the reaction mixture.

The ^{32}P end-labeled DNA fragments were separated from one another by electrophoresis through agarose (Hohn et al., 1980) or polyacrylamide (Peacock and Dingman, 1967) gels. Fragments were recovered from agarose gels by the method of Vogelstein and Gillespie (1979). Elution from polyacrylamide gels was by agitation of the crushed gel band overnight at 37°C in 500 mM NaCl, 50 mM Tris (pH 7.9). Soluble polyacrylamide was eliminated by adsorption of the DNA to a small DEAE-cellulose column and subsequent elution at high salt concentration. Fragments were concentrated by ethanol precipitation with 10 μg carrier tRNA.

The 5′ labeled ends of complementary strands were separated by digestion with a second restriction enzyme or, more frequently, by strand separation. Strand separation was carried out by heating the DNA fragment at 90°C for 2 min in the presence of 20% dimethylsulfoxide. The sample was then quick-chilled and immediately loaded on a 5 or an 8% (depending on fragment size) polyacrylamide gel (Szalay, Grohmann and Sinsheimer, 1977).

Singly end labeled DNA fragments were eluted from crushed polyacrylamide gel bands as described above except that the DEAE-cellulose column step was omitted. After addition of 10 μg carrier tRNA, the fragments were ethanol-precipitated, resuspended in 60 mM sodium acetate and reprecipitated, and the precipitate was washed with 90% ethanol. Finally, the precipitate was dried in vacuo and resuspended in a small volume of distilled water for the sequencing reactions.

For labeling 3′ ends, about 2 μg of purified CaMV DNA restriction fragment were dissolved in 25 μl 0.1 M potassium cacodylate (pH 7.6), 1 mM $CaCl_2$, 0.2 mM dithiothreitol (Roychoudhury, Jay and Wu, 1976), and incubated at 37°C for 10 min with 2 units of calf thymus terminal deoxynucleotidyl transferase (PL Biochemicals) and 50 μCi

of α-^{32}P-ATP (New England Nuclear). An additional 2 units of enzyme were added and incubation was continued for 30 min, at which time the mixture was supplemented with 100 nmole ATP and incubated for 10 min more. The DNA was ethanol-precipitated with 5 μg carrier tRNA and the precipitate, collected by centrifugation, was dissolved in 100 μl 1 M piperidine. After heating to 90°C for 30 min, piperidine was eliminated by three cycles of lyophilization. The lyophilized DNA, taken up in distilled water, was subjected to strand separation as described above.

Base-specific chemical cleavage reactions were carried out according to the methods of Maxam and Gilbert (1977), using the methylation reaction for G, depurination for G + A and hydrazine reactions for C + T and C. The cleavage products were fractionated on 8 or 20% 0.5 mm thick polyacrylamide sequencing gels operated at high voltage (Sanger and Coulson, 1978). Autoradiography was performed at −70°C with phosphotungstate intensifying screens. Approximately 75% of the genome was sequenced on both strands with special attention being devoted to those portions of the coding region where there is a shift in reading frame (see Results and Discussion). A computer program was used to search for restriction enzyme sites and sequence overlaps in the course of construction of the sequence, and another program was written to search the finished sequence for potential coding regions and for sequences resembling splice junctions (Lerner et al., 1980). Details concerning derivation of the sequence and photographs of sequencing gels are available upon request.

Acknowledgments

The authors would like to thank G. Lebeurier and A. Lesot for good advice on restriction enzymes and other matters, and G. Daney de Marcillac for computer programming. This work was supported by the Centre National de la Recherche Scientifique.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 2, 1980

References

Al Ani, R., Pfeiffer, P. and Lebeurier, G. (1979). The structure of cauliflower mosaic virus. II. Identity and location of the viral polypeptides. Virology *93*, 188–197.

Al Ani, R., Pfeiffer, P., Whitechurch, O., Lesot, A., Lebeurier, G. and Hirth, L. (1980). A virus-specified protein produced upon infection by cauliflower mosaic virus. Ann. Virol. (Inst. Pasteur) *131E*, 33–53.

Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980). The ovalbumin gene—sequence of putative control regions. Nucl. Acids Res. *8*, 127–142.

Brunt, A. A., Barton, R. J., Tremaine, J. H. and Stace-Smith, R. (1975). The composition of cauliflower mosaic virus protein. J. Gen. Virol. *27*, 101–106.

Civerolo, E. L. and Lawson, R. H. (1978). Topological forms of cauliflower mosaic virus nucleic acid. Phytopathology *61*, 188–193.

Hohn, T., Hohn, B., Lesot, A. and Lebeurier, G. (1980). Restriction map of original and cloned cauliflower mosaic virus DNA. Gene, in press.

Howell, S. H. and Hull, R. (1978). Replication of cauliflower mosaic virus and transcription of its genome in turnip leaf protoplasts. Virology *86*, 468–481.

Howell, S. H., Odell, J. T. and Dudley, K. R. (1979). Expression of the cauliflower mosaic virus genome in turnips (Brassica rapa). In Genome Organization and Expression in Plants, C. J. Leaver, ed. (New York: Plenum Press).

Hull, R. (1979a). The DNA of DNA plant viruses. In Nucleic Acids in Plants, Vol. II, T. Hall and J. W. Davies, eds. (Cleveland: CRC Press).

Hull, R. (1979b). Structure of the cauliflower mosaic virus genome. III. Restriction endonuclease mapping of thirty-three isolates. Virology *100*, 76–90.

Hull, R. and Shepherd, R. J. (1976). The coat proteins of cauliflower mosaic virus. Virology *70*, 217–220.

Hull, R. and Shepherd, R. J. (1977). The structure of the cauliflower mosaic virus genome. Virology *79*, 216–230.

Hull, R. and Howell, S. H. (1978). Structure of the cauliflower mosaic virus genome. II. Variation in DNA structure and sequence between isolates. Virology *86*, 482–493.

Hull, R., Shepherd, R. J. and Harvey, J. D. (1976). Cauliflower mosaic virus: an improved purification procedure and some properties of the virus particles. J. Gen. Virol. *31*, 93–100.

Hull, R., Covey, S. N., Stanley, J. and Davies, J. W. (1979). The polarity of the cauliflower mosaic virus genome. Nucl. Acids Res. *7*, 669–677.

Kelly, D. C., Cooper, J. and Walkey, D. G. A. (1974). Cauliflower mosaic virus structural proteins. Microbios *10*, 239–245.

Lebeurier, G., Whitechurch, O., Lesot, A. and Hirth, L. (1978). Physical map of DNA from a new cauliflower mosaic virus strain. Gene *4*, 213–226.

Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. and Steitz, J. A. (1980). Are small snRNPs involved in splicing? Nature *283*, 220–224.

Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci. USA *74*, 560–564.

Paterson, B. M., Roberts, B. E. and Kuff, E. L. (1977). Structural gene identification and mapping by DNA-mRNA hybrid arrested cell-free translation. Proc. Nat. Acad. Sci. USA *74*, 4370–4374.

Peacock, A. C. and Dingman, C. W. (1967). Resolution of multiple ribonucleic acid species by polyacrylamide gel electrophoresis. Biochemistry *6*, 1818–1827.

Philipson, L. (1979). Adenovirus proteins and their messenger RNAs. Adv. Virus Res. *25*, 357–405.

Proudfoot, N. J. and Brownlee, G. G. (1976). 3′ noncoding region sequences in eukaryotic messenger RNA. Nature *263*, 211–214.

Roychoudhury, R., Jay, E. and Wu, R. (1976). Terminal labeling and addition of homopolymer tracts to duplex DNA fragments by terminal deoxynucleotidyl transferase. Nucl. Acids Res. *3*, 863–877.

Russell, G. J., Follet, E. A. C., Subak-Sharpe, J. H. and Harrison, B. D. (1971). The double-stranded DNA of cauliflower mosaic virus. J. Gen. Virol. *11*, 129–138.

Sanger, F. and Coulson, A. R. (1978). The use of thin acrylamide gels for DNA sequencing. FEBS Letters *87*, 107–110.

Sanger, F., Air, G. M., Barell, B. G., Brown, N. L., Coulson, A. R., Fiddles, J. C., Hutchison, C. A., Slocomb, P. M. and Smith, M. (1977). Nucleotide sequence of bacteriophage ϕX174 DNA. Nature *265*, 687–695.

Shepherd, R. J. (1979). DNA plant viruses. Ann. Rev. Plant Physiol. *30*, 405–423.

Shepherd, R. J. and Wakeman, R. J. (1971). Observation on the size and morphology of cauliflower mosaic virus deoxyribonucleic acid. Phytopathology *61*, 188–193.

Shepherd, R. J., Bruening, G. E. and Wakeman, R. J. (1970). Double-stranded DNA from cauliflower mosaic virus. Virology *41*, 339–347.

Szalay, A. A., Grohmann, K. and Sinsheimer, R. L. (1977). Separation of the complementary strands of DNA fragments on polyacrylamide gels. Nucl. Acids Res. *4*, 1569–1578.

Tezuka, N. and Taniguchi, T. (1972). Structural protein of cauliflower mosaic virus. Virology *48*, 297–299.

Vogelstein, B. and Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. Proc. Nat. Acad. Sci. USA *76*, 615–619.

Volovitch, M., Drugeon, G. and Yot, P. (1978). Studies on the single-stranded discontinuities of the cauliflower mosaic virus genome. Nucl. Acids Res. *5*, 2913–2925.

Volovitch, M., Drugeon, G., Dumas, J. P., Haenni, A. L. and Yot, P. (1979). A restriction map of cauliflower mosaic virus DNA (strain PV 147). Eur. J. Biochem. *100*, 245–255.

plant JOURNAL

FOR CELL AND MOLECULAR BIOLOGY

The Plant Journal (1992) 2(2), 233–239

Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene

Helmut Bäumlein*, Istvan Nagy†, Raimundo Villarroel[1], Dirk Inzé[1] and Ulrich Wobus
Institut für Genetik und Kulturpflanzenforschung, Corrensstraße 3, Gatersleben, O-4325, Germany, and [1]Laboratorium voor Genetika, Rijksuniversiteit Gent, Ledeganckstraat 35, Gent, B-9000, Belgium

Summary

A 2.4 kb fragment containing the 5′-flanking region and the 5′-noncoding sequence of the *Vicia faba* legumin gene *LeB4* mediates high level seed-specific expr ssion in transgenic tobacco plants. Deleted derivatives of this legumin upstream sequence were fus d to the *npt-II* reporter gene to determine the tissue-specific activity of the chimeric constructs in stably transformed tobacco plants. The results indicate the presence of positive regulatory, enhancer-lik *cis* elements within 566 bp of the upstream sequence. Most importantly, however, these elements are only fully functional in conjunction with the core motif CATGCATG of the legumin box around position −95, since destruction of the motif by a 6 bp deletion in an otherwise intact 2.4 kb upstream sequence drastically reduces expression in seeds. At the same tim , low level expression in leaves is observed. The ccurrence of similar CATGCATG consensus *cis* elements with alternating purine and pyrimidine base pairs in front of several other plant genes suggests a functional role of the motif in a wider range of plant promoters.

Introduction

Spatially and temporarily regulated gene expression programmes are the basis for development and morphology. The strictly seed-specific and development-dependent expression of seed storage protein genes provides a suitable experimental system to study differential gene activation in plants.

It is generally accepted that the seed specificity of storage protein gene expression is primarily regulated at the transcriptional level, although post-transcriptional processes can modulate the final amount of translational products widely (Goldberg *et al.*, 1989). Current ideas imply complex interactions between specific *trans*-acting transcription factors with their *cis*-acting target DNA sequences as the principal mechanism for transcription regulation. Several DNA fragments derived from the 5′-flanking regions of different seed protein genes have been shown to bind defined nuclear protein factors (Allen *et al.*, 1989; Bustos *et al.*, 1989; Chen *et al.*, 1988; Jofuku *et al.*, 1987; Jordano *et al.*, 1989). However, in most cases a causal relationship connecting *trans* factor binding with regulated promoter activity has not been demonstrated. The availability of extensive sequence data from 5′ flanking regions of storage protein genes isolated from several different species has prompted the search for conserved sequence motifs, assuming that these elements might be involved in *trans* factor binding and therefore in the regulation of seed protein gene expression. Thus several sequence conservative, putative regulatory DNA elements have been identified (for review see Okamuro and Goldberg, 1989); among them the legume 12S globulin gene-specific legumin box (Bäumlein *et al.*, 1986) with the internal, highly conserved RY core motif CATGCATG (Dickinson *et al.*, 1988).

Recently we have shown that about 1.2 kb of the legumin B4 (*LeB4*) gene upstream sequence is sufficient for strong seed-specific activity and that deletion derivatives with only 193 bp and 91 bp of upstream sequence are approximately 10 times less active (Bäumlein *et al.*, 1991a). For a more precise localization of the *cis* elements which might be responsible for this reduction in activity we have constructed and analysed a series of new deletions.

In this paper we present data extending our knowledge of functionally important DNA sequences in the 5′-flanking region of the gene *LeB4*. In particular, we demonstrate that strong legumin promoter activity and probably also strict tissue specificity depend on the integrity of the short conserved CATGCATG sequence motif within the legumin box.

Results

Delineation of cis*-acting elements by 5′ deletion analysis*

Earlier experimental data (Bäumlein *et al.*, 1991a) demonstrate the presence of functionally important elements in

Received 5 April 1991; revised 26 November 1991.
*For correspondence (fax +37 45592 280).
†Present address: Agricultural Biotechnology Centre, H-2101 Gödöllö, Hungary.

the legumin *LeB4* gene upstream sequence between about 1200 bp (*Cla*I site) and 193 bp (*Eco*RV site) in front of the transcription start site. For a more precise characterization of those elements we have analysed the effect of progressive and internal deletions within about 1 kb of the *LeB4* upstream sequence (see Figure 1) on NPT-II reporter enzyme levels in seeds of stably transformed tobacco plants. As a first approximation we interprete the changes in enzyme activity as a reflection of changes in promoter strength.

As shown in Figure 2 the *LeB4* upstream sequence can be deleted to position −701 without an obvious loss of NPT-II activity. The average enzyme activity seems to drop when the sequence between −701 and −566 is removed. However, this transition is not statistically significant, and neither is the increase in activity between construct −844 and −701. A significant (at the 5% level) reduction in expression level can be detected when the promoter is shortened to −471 bp. The 95 bp sequence between −566 and −471 is AT-rich (73%) and includes the motif ATTAATT which partly satisfies the ATT A/T AAT consensus rule (Jofuku *et al.*, 1987). The *Ppu*MI site at position −492 used for the construction of the two internal deletions PC and PR (see Figure 1) is also located within this sequence. This restriction site overlaps a so-called GC element present in all legumin gene upstream sequences surveyed (Rerie, 1989; Rerie *et al.*, 1991).

Another extremely AT-rich (82%) region was removed to obtain construct −407. The enzyme levels produced by this construct are on average less than 10% compared to those produced by constructs −701 or −844. Construct −333 lacks part of a DNA motif with a 20 out of 25 bp homology (see Figure 1) to a promoter sequence of the mainly seed-specifically expressed *USP* gene of *Vicia faba* (Bäumlein *et al.*, 1991b) with no obvious effect. Another significant (at the 1% level) reduction in the expression level is shown by construct −232 in comparison to construct −279. The removed sequence does not show any obvious peculiarity apart from an 11-bp purine stretch.

The question of whether a minimal promoter completely lacking the conservative legumin box is still functional was addressed by the analysis of construct −68. Construct −68 leads to significantly (at the 1% level) reduced but still measurable NPT-II activities in comparison to construct −151. The sequence between position −151 and −68 bp includes the total legumin box and an imperfect direct repeat (TGTCACACACGTtcTGTCACACGT) between position −83 and −60 with similarity both to a motif reported to be present in front of several plant genes (Memelink *et al.*, 1987) and to the CACA motif often found in the upstream regions of seed protein genes (Okamuro and Goldberg, 1989). The effects of even shorter promoter constructs on NPT-II activities in seeds have been compared in a separate experimental series. A 45-bp long

(a)

```
                ClaI(ca.-1200).............ca.50bp.........
AAAAAAGTAATGCAGACAAAAAGTGGAAAAGACAATCTGACTGTACATAAGAAATTTCCA
ATTTTTGAAATTTTTTTATAATTATCAGAAATTTTAAAATTTCCGATAAAAACATACATC
TATAGATCGAAAATTTCAAATTTCTAGTACTTTCAAATTTCTTGCAGTAAAAGTTGTAAT
TTTTTAAAAATTTACGATAATTTACAGTATTTAAAAAAAAAATCCAATCTTAAATAAAGG
GTATAAGAATAAAAGCACTCATGTGGAGTGGCAGGTTTCGTCACACCCTAAGAACATCCC
          *-844
TAAATACACCACATATGTATAAGTATTAAGTGATTGATGTTAAGTGAAACGAAAATATTT

ATATGTGAAATTTAATATTCAGCTTACTTGATTAAACTCCATAGTGACCCAATAAGTGCT
                                                  *-701
AACTTTTACTGTCTTTACCTTTAAATGTTATATTGATTTATTTATGCATTTCTTTTTCCT

GCATCTCAATAGTATATAGGGTATCAAATAGTGATTATCCAAACTTAAATAAGTTAGAGG

AAACACCAAGATATGCCATATACTCTCAAATTTGACACTATGATTCAAAGTTGCACTTGC
*-566
ATAAAACTTATTAATTCAATAGTAAAACCAAACTTGTGCGTGATACAGTTAAAATGACTA
             *PpuMI(-492)          *-471
AACTACTAATTAAGGTCCCTCCCATTAGTAAATAAGTTATTTTTTTAGAAAAAGAAAATA
                                            *-407
ATAAAAAGAATGACGAGTCTATCTAAATCATATTAACAAGTAATACATATTGATTCATTC
                                                          *-333
GATGGAGGAGGCCAATAATTGTAGCAAACAAGCAGTGCCGAGGTTAATATATGCTCAAGA
       ::::AT::::::G:::C::C::::: USP homology
                                                    *-279
CAGTAAATAATCTAAATGAATTAAGACAGTGATTTGCAAAGAGTAGATGCAGGAGAGA
                                     *-232
ACTAAAGATTTGCTGCTACACGTATATAAGAATAGCAACAGATATTCATTCTGTCTCTTT
            *EcoRV(-193)                                   *-151
GTGGAATATGGATATCTACTAATCATCATCTATCTGTGAAGAATAAAAGAAGCGGCCACA
                                                            *SphI(-91)
AGCGCAGCGTCGCACATATGATGTGTATCAAATTAGGACTCCATAGCCATGCATGCTGAA
                  *-68                           *-45
GAATGTCACACACGTTCTGTCACACGTGTTACTCTCTCACTGTTCTCCTCTTCCTATAAA
            *-14          *cap                 *+20
TCACCGCGCCACAGCTTCTCCACTTCACCACTTCACCACTTCACTCACAATCCTTCATTA
                              linker      Met Trp
GTTGTTTACTATCACAGTCACAcggatcgatctgatcATG TGG-npt II gene
```

(b)

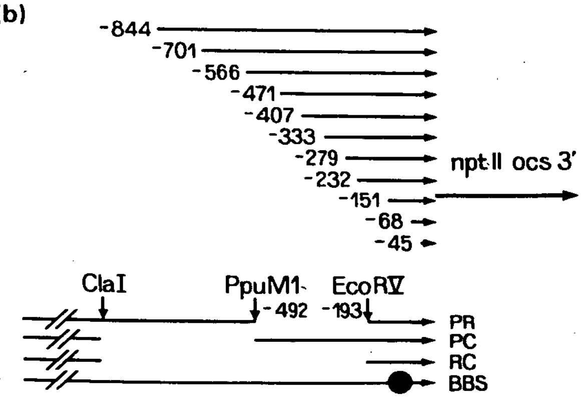


Figure 1. Sequence of the 5′-flanking region of *LeB4* and structure of the *LeB4* promoter deletion constructs.

(a) Sequence of the 5′-flanking region of the legumin gene *LeB4* fused to the *npt-II* coding region in the Ti plasmid pGV180 by a linker region. The start points for deletion derivatives are indicated by * above the sequence and the position number is supplemented by a restriction enzyme symbol in case the respective site was used to create the deletion construct. The *Cla*I site indicated at the top has been mapped to about 50 bp upstream of the given sequence but not sequenced itself. Sequence motifs discussed in the text are marked by underlining and the CATGCATG motif within the legumin box is denoted by italics. The linker region between the last nucleotide of the *LeB4* 5′-noncoding region and the first two codons of the *npt-II* reporter gene are printed in lower case letters. The sequence between positions −689 and +56 has already been published by Bäumlein *et al.* (1986).

(b) Schematic structure of the *LeB4* promoter deletion constructs used in this study. The arrow at the right labelled nptII ocs 3′ symbolizes the neomycinphosphotransferase-II reporter gene terminated by the polyadenylation region of the octopine synthase gene. The other arrows represent *LeB4* sequences upstream of the *npt-II* fusion point indicated in (a) and labelled by the respective deletion end-points. Constructs denoted PR, PC, RC and BBS were created by deletions within the total 2.4 kb *LeB4* upstream region by removing the indicated restriction fragments or, in the case of the BBS construct, by deleting 6 bp of the legumin box core motif CATGCATG, as specified in Figure 3.

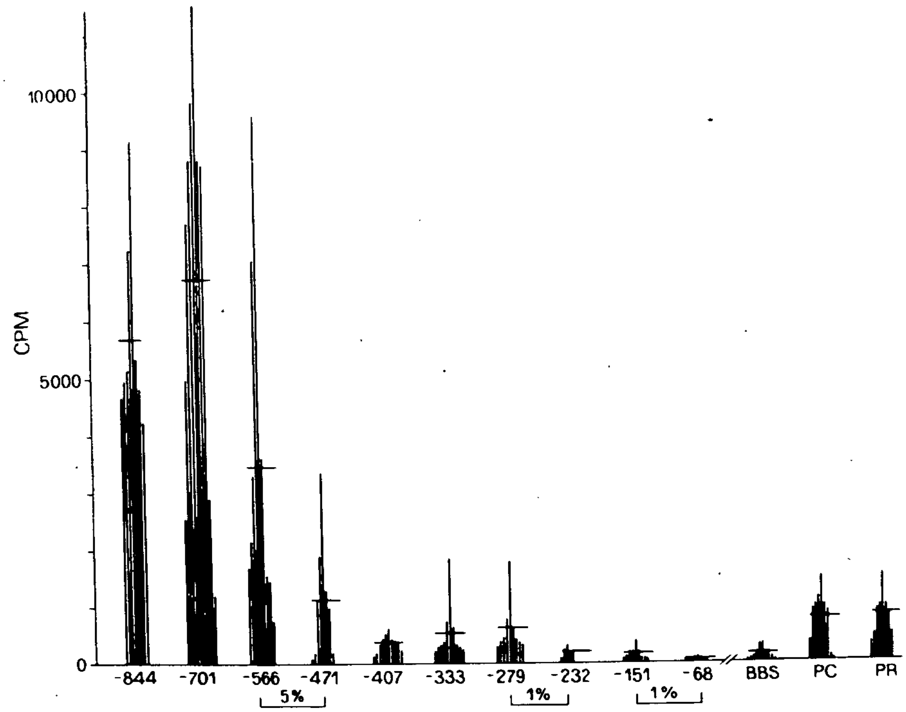


Figure 2. NPT-II activity levels measured by the NPT-II gel assay in mature seeds of independent tobacco plant transformants.
The average value from all plants transformed with the constructs indicated below the columns is denoted by a horizontal line. The constructs are defined in Figure 1. Statistically significant differences at the 1% or 5% significance level between consecutive constructs are indicated by brackets. Fifty micrograms of protein were used in each assay. To keep experimental variability as low as possible, all values given were estimated in a single slot-blot experiment.

promoter (construct −45) still causes low NPT-II activity comparable to that of construct −68. Only the removal of the TATA box in construct −14 completely extinguishes the promoter activity. In addition, a cap site deletion (construct +20) is also inactive, as expected (data not shown).

The progressively shortened promoter constructs described above necessarily change the spatial relationship between the transcription start site and the flanking vector-derived sequences as well as the sequences adjacent to the genomic integration site. To reduce the potential influence of spatial changes we have created and analysed deletions within the 2.4 kb 5′ flanking region (see Figure 1). As shown in Figure 2, both the PR and PC constructs show strongly decreased expression. The low activity of PR confirms the data obtained with the progressive deletion constructs −471, −407, −333, −279 and −232 and demonstrates the necessity of sequence elements between −492 (*Ppu*MI site) and −193 (*Eco*RV site) for optimal promoter function. Moreover, the reduced expression of the PC construct indicates that additional sequence elements at or upstream of the *Ppu*MI site quantitatively affect the expression of the legumin promoter. Considering that sequences upstream of position −566 can be deleted without a significant effect (Figure 2), we conclude that those additional sequence elements are localized closely upstream of or even overlapping the *Ppu*MI site at −492.

The legumin box core motif CATGCATG is essential for seed-specific promoter activity

Assuming that sequence conservation is an indication of functional importance, it has been suggested that the legumin box and its core motif CATGCATG are crucial for legumin gene expression (Bäumlein *et al.*, 1986; Dickinson *et al.*, 1988). To test this hypothesis experimentally we have used a suitable unique *Sph*I site overlapping the CATGCATG core element of the legumin box to specifically remove 6 bp out of the 8 bp core motif (see Figure 3) in the 2.4 kb *LeB4* upstream sequence (BBS deletion). All of the 10 individual transformants analysed show low NPT-II activity in mature seeds, comparable in intensity to the enzyme levels caused by construct −151 (Figure 2). Surprisingly, seven out of the 10 plants transformed with the BBS construct also showed low NPT-II activity in leaves. Examples are given in Figure 4. In contrast, leaf activity is not found in plants carrying constructs with at least 700 bp proximal to the *LeB4* transcription start site (data not shown). To exclude additional unintended changes within the mutated fragment as a cause for the low and tissue-specifically relaxed NPT-II levels, we have confirmed the overall integrity of all BBS constructs by Southern hybridization. Moreover, the removal of the former *Sph*I site was proven by the resistance to *Sph*I treatment of a legumin box containing PCR fragment amplified from genomic DNA of BBS-transformed tobacco

GmGly	TCCATAGC***CATGCATA***CTGAAGAATG
PsLegA	TCCATAGC***CATGCAAG***CTGCAGAATG
PsLegJ	TCCATAGC***CATGCATG***CTGAACAATG
VfLegB	TCCATAGC***CATGCATG***CTGAAGAATG
BBS	TCCATAGCCA******CTGAAGAATG
GmβCG	AGC***CATGCA***
	C***CATGCATG***
Asglo5	T***CAT-CATG***
ZmC1	TC***CATGCATG***CAC
	TG***CATGCATG***CAC
ZmRAB17	TCCACT***CATGCAT***
	CT***CATGCATG***CCC
OsRAB16	TCCACC***CATGCCG***
TsEm	TG***CATGCATG***CAA
Gmaux22	***CATGCAT***
SV40	AAG***CATGCATCT***C
	AAG***TATGCA***

Figure 3. CATGCATG-like motifs are present in front of several plant genes as well as in the *SphI* element of the SV40 enhancer.
Abbreviations: GmGly, *Glycine max*, glycinin gene; PsLegA, *Pisum sativum* legumin A gene; PsLegJ, *P. sativum*, legumin J and K genes (Thompson *et al.*, 1991); VfLegB, *Vicia faba*, legumin B gene (Bäumlein *et al.*, 1986); BBS, 6 bp deletion within the legumin box (this paper); GmβCG, *G. max*, β-conglycinin gene (Harada *et al.*, 1989); Asglo5, *Avena sativa*, 12S globulin gene (Schubert *et al.*, 1990); ZmC1, *Zea mays*, *C1* regulator gene of anthocyan synthesis (Paz-Ares *et al.*, 1987); ZmRAB17, *Z. mays*, abscisic acid-induced gene (McCarty, personal communication); OsRAB16, *Oryza sativa*, abscisic acid-responsive gene (Mundy *et al.*, 1990); TsEm, *Triticum aestivum*, abscisic acid-induced wheat gene (McCarty, personal communication); Gmaux22, *G. max*, auxin-regulated gene (Ainley *et al.*, 1988); SV40, *SphI* element in the simian virus 40 enhancer (Zenke *et al.*, 1986).

plants. Thus experimental data clearly demonstrate that the destruction of the conservative RY motif CATGCATG within the 2.4 kb upstream region strongly disturbs the function of the legumin B4 promoter.

The AT-rich RC fragment enhances the activity of a truncated foreign promoter

Earlier experimental data (Bäumlein *et al.*, 1991a) demonstrate that the generally AT-rich region between positions

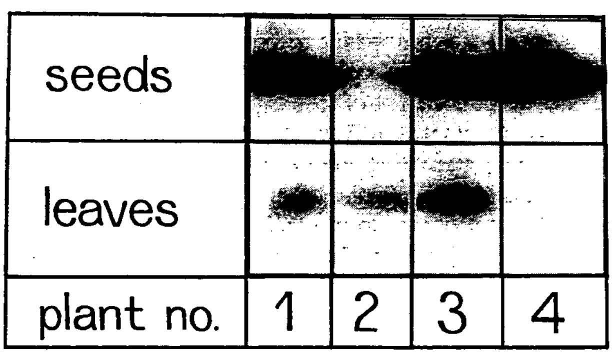


Figure 4. NPT-II activity in seeds and leaves of four individual plants transformed with the BBS construct.
All extracts containing 50 μg of protein were assayed on the same gel and the autoradiogram exposed for 3 days. Note that there is no obvious correlation in activity between seeds and leaves in each construct.

−1200 (*ClaI* site) and −193 (*EcoRV* site) exhibit a clear-cut quantitative effect on the basic *LeB4* promoter. The same promoter region can also co-operate with the truncated *nos* promoter in the Ti-plasmid-derived vector pGV300 which contains 148 bp of upstream sequence still including the b, a, z and reversed b element configuration (Ebert *et al.*, 1987). As shown in Table 1, the RC fragment in constructs RCD+ and RCD− enhances the activity of the truncated *nos* promoter in leaves more than 25-fold, independent of its orientation. Surprisingly, in seed tissue, the enhancing effect is only two- to fourfold; the difference in NPT-II activity between the two orientations is not statistically significant. In contrast to the RC fragment, a legumin box containing *MboII* fragment (positions −155 to −77 in Figure 1) in constructs LBL+ and LBL− does not (or only weakly) interact with the *nos* promoter in seeds, whereas in leaves the reverse but not the natural orientation increases *nos* promoter activity about sixfold (Table 1).

Discussion

Several upstream elements quantitatively influence the legumin promoter activity

The functional analysis in transgenic tobacco plants of a series of deletions covering about 1 kb of *LeB4* gene upstream sequences specifies further earlier conclusions about *LeB4* promoter regulation (Bäumlein *et al.*, 1991a). As shown in Figure 2 there is a highly significant decrease in activity when constructs −701 and −471 are compared. The data suggest that the region distal of −566, but certainly distal of −701, is of little importance for high promoter activity in contrast to the region downstream of bp −566. This region up to bp −407 is rich in AT base pairs. Similar AT-rich sequences have been described as being involved in the regulation of genes coding for seed and other plant proteins. These sequences preferentially interact with high mobility group (HMG) proteins which seemingly recognize certain structural features instead of specific primary sequences (reviewed by Weising and Kahl, 1991). Within these AT-rich sequences lies the *Ppu* MI site-overlapping, evolutionary conserved GC element AAGGTCCCT (Rerie, 1989; Rerie *et al.*, 1991). We take its sequence conservation and the reduced NPT-II activity of the PR and PC constructs (Figure 2), in which either the 5′ or the 3′ part of the *PpuMI* site are removed, as an indication of the functional importance of the GC element.

Another significant transition in activity, although already at a low level, occurs when the fragment between positions −151 and −68, containing the legumin box, is removed (Figure 2). Whereas the BBS deletion clearly reveals the importance of the legumin box core motif CATGCATG (see below) the role, if any, of the additional box sequences

Table 1. Effect of *LeB4* upstream region fragments on a truncated *nos* promoter in transgenic tobacco plants

	Mean (± SEM) value of NPT-II activity (c.p.m.)		Total number of plants
	Seed	Leaf	
pGV300	120 ± 12	343 ± 111	9
LB14	179 ± 17*	344 ± 46	2
LB11	139 ± 11	2038 ± 743**	10
RCD37	270 ± 64**	9790 ± 3975**	4
RCD2	500 ± 195**	9202 ± 3298**	4

pGV300 is the control Ti plasmid described in Experimental procedures containing the truncated *nos* promoter fused to the *npt-II* gene. We fused either the legumin box containing promoter fragment −156 to −77 in the natural (LBL+) or the inverse orientation (LBL−), or the RC fragment (−1200 to −193), again in either the natural (RCD+) or the inverse orientation (RCD−) in front of the truncated promoter. Fifty micrograms of protein extracted from leaves or mature seeds of a total of 29 transgenic plants were analysed by the NPT-II gel assay. Significant difference, at the *5% or **1% level, between a given construct and the control pGV300 in either seeds or leaves as calculated by the Mann–Whitney U test.

around the core motif remains undefined. The low but significant NPT-II activities in seeds of plants transformed with the *LeB4* promoter constructs −232, −151 and −68 are in contrast to results reported by Shirsat *et al.* (1989) and Rerie *et al.* (1991). These authors tested promoter deletions of the pea legumin gene *Leg1* by estimating Leg1 protein levels in transgenic tobacco seeds and were unable to detect any expression when upstream sequences of only 97 bp, 124 bp and 237 bp control legumin expression. The difference between these and our results may be explained by the lower detection sensitivity of the immunological technique used by Shirsat *et al.* (1989) and Rerie *et al.* (1991), although differences due to the constructs used (intact gene versus chimeric gene) cannot be excluded. Presently we cannot fully explain the results of *nos* promoter stimulation by *LeB4* promoter fragments (see Table 1) but we initially conclude that (i) the AT-rich RC fragment contains sequences which meet the criteria for enhancers (Müller *et al.*, 1988) in stimulating the foreign minimal *nos* promoter in an orientation-independent manner, especially in leaves, and (ii) there is no element within the RC fragment acting as a *seed-specific* enhancer in the given construct.

The CATGCATG motif – a key element of the legumin gene promoter

The sequence motif CATGCATG is conserved among legume seed protein genes (Dickinson *et al.*, 1988) and is part of the 28 bp legumin box found in front of genes coding for 12S legume seed globulins (Bäumlein *et al.*, 1986). The exclusive deletion of 6 out of 8 bp of the CATGCATG motif within the 2.4 kb *LeB4* upstream sequence in front of the *npt-II* reporter gene leads to a dramatic reduction of NPT-II enzyme levels (see Figure 2). However, since similar reductions are caused by progressive deletions (−232, −151) leaving the CATGCATG motif intact, we conclude that this motif is necessary but not sufficient for optimal promoter function. These data also explain why we were unable to demonstrate the functional importance of the legumin box using progressive deletions only (Bäumlein *et al.*, 1991a) and imply that the legumin box core element CATGCATG can only function properly in co-operation with additional upstream elements.

Destruction of the CATGCATG motif also causes low NPT-II activity in leaves of BBS plants (Figure 4). Such leaf activity has been already observed in plants carrying the RC deletion construct (see Figure 1) as well as −193 and −91 constructs (Wobus *et al.*, 1989). Relaxed tissue specificity was also reported for shortened patatin-1 promoter constructs (Jefferson *et al.*, 1990) and for a truncated anonymous root-specific promoter (Koncz *et al.*, 1989). We favour the idea that the *LeB4* promoter loses its tissue specificity when the promoter is turned down by the removal or destruction of important *cis* elements. However, we have still not rigorously excluded other explanations, such as an unknown role of the *npt-II* coding sequence, as described for mammalian cells by Artelt *et al.* (1991).

The CATGCATG motif also occurs in other plant gene promoters

Although originally described as an element specific for legume seed protein genes, here we suggest that the CATGCATG motif acts as a functional module in a wider range of plant promoters. Figure 3 shows its physical

presence within the upstream regulatory sequences of several plant genes as well as the SV40 *Sph*I enhancer motif. At least for the maize *C1* gene it was shown that the CATGCATG sequence is crucial for its regulation by the *viviparous* gene product Vp1 (McCarty and Carson, 1991; McCarty, personal communication). We presently favour the idea that either a CATGCATG-binding transcription factor or a structural peculiarity due to the alteration of purine and pyrimidine bases, or both, are involved in the integration of a functional transcription complex in seed tissue.

Experimental procedures

Plasmid constructs

Standard cloning, construction and sequencing techniques have been performed following the guidelines given in Ausubel *et al.* (1987) and Sambrook *et al.* (1989). The starting point for the generation of progressively deleted promoter fragments was the plasmid p4/12BB, described previously (Bäumlein *et al.*, 1991a). p4/12BB contains, beside pUC18 vector sequences, a 2.4 kb upstream region with unique restriction sites for *Cla*I (around −1200), *Ppu*MI (−492), *Eco*RV (−193) and *Sph*I (−91) plus the complete 56 bp 5′-untranslated region of gene *LeB4*. The whole fragment is flanked by an upstream *Eco*RI/*Bgl*II/*Sma*I linker sequence and a downstream *Bam*HI site. After cleavage at the *Cla*I site, p4/12BB was partially digested with *Bal* 31, re-cut with *Sma*I and recircularized. The deletion end-points were determined by the Sanger sequencing technique.

To create the BBS construct, plasmid p4/12BB was cut at the *Sph*I site overlapping the CATGCATG sequence motif, treated with T4 DNA polymerase to resect the 3′ protruding ends, and recircularized. The cloned products were sequenced to analyse the extent of the deletion.

The *Ppu*MI site, dividing the *Eco*RV/*Cla*I fragment (RC) into a distal and a proximal part was used to create the two internal deletions, PC (removing the distal *Cla*I/*Ppu*MI fragment) and PR (removing the proximal *Ppu*MI/*Eco*RV fragment).

The deleted promoter fragments were isolated as *Bgl*II/*Bam*HI fragments and cloned in the right orientation into the *Bgl*II site of the intermediate vector pGV180 containing a promoterless *npt-II* gene (see Bäumlein *et al.*, 1991a). Another strategy was applied to create the three promoter constructs −45, −14 and +20. In this case, the unique *Cla*I site of the plasmid pGV180/legP FL (Bäumlein *et al.*, 1991a), containing the same *LeB4* sequences as plasmid p4/12BB described above, was used as the start point for the partial *Bal* 31 digestion. Again the digestion products were cut with *Sma*I to remove the upstream sequences, gel-purified, recircularized, transformed and the deletion end-points determined by sequence analysis.

To test the influence of several *LeB4* promoter fragments on a truncated foreign promoter, we used the enhancer trap vector pGV300, originally designed by Allan Caplan, Rijksuniversiteit Gent. In this plasmid, which was derived from the pGV180 vector (Bäumlein *et al.*, 1991a; Herman *et al.*, 1986) the *npt*-II reporter gene is driven by a truncated *nos* promoter. Using a suitable *Sst*II site the *nos* promoter was shortened to a length of 148 bp, still including the b, a, z and reversed b sequence elements described to be important for (albeit reduced) promoter activity (Ebert *et al.*, 1987). Both the *Eco*RV/*Cla*I fragment (RC) and the legumin box-containing *Mbo*II fragment (LBL) spanning from position −156 to −77 have been cloned in either orientation in front of this truncated *nos* promoter.

Plant transformation

The intermediate plasmids were transferred into the *Agrobacterium* strain pGV2260 by triparental mating and used for leaf disc transformation of *Nicotiana tabacum* cv. Havana as described previously (Bäumlein *et al.*, 1991a). The integrity of all constructs was checked both in *Agrobacterium* and in the plants using Southern hybridization and PCR techniques.

NPT-II assays

NPT-II activity was detected in 100 mg of tissue. Equal amounts of protein determined by the Bradford assay, were assayed for NPT-II activity either by the gel test (Reiss *et al.*, 1984) or the dot technique (Platt and Yang, 1987). For quantification, the radioactivity of cut filter spots was counted. Seed NPT-II activity was determined from each individual transformant and the grouped values compared by the Mann–Whitney U test. In another experiment, equal amounts of seeds (100 mg each) of all transformants harbouring the same construct were mixed, extracted and analysed on a single gel. The principal results (not shown) did not deviate from those shown in Figure 2 for individual transformants.

Acknowledgements

We thank Andreas Czihal, Elke Liemann, Angela Stegmann, Hannelore Weiβ and Thea Tölle for excellent technical assistance, Drs Allan Caplan and Diana Jofuku for helpful discussions, Dr Ute Heim for assistance with DNA sequencing, and Dr Kirk Schnorr for critical reading of the manuscript. UW acknowledges the support given by the 'Fond der Chemischen Industrie'. HB would also like to thank Professor Marc Van Montagu for his hospitality and advice.

References

Ainley, W.M., Walker, J.C., Nagao, R.T. and Key, J.L. (1988) Sequence and characterization of two auxin regulated genes from soybean. *J. Biol. Chem.* **263**, 10658–10666.

Allen, R.D., Bernier, F., Lessard, P.A. and Beachy, R.N. (1989) Nuclear factors interact with a soybean β-conglycinin enhancer. *Plant Cell*, **1**, 623–631.

Artelt, P., Grannemann, R., Stocking, C., Friel, J., Bartsch, J. and Hauser, H. (1991) The prokaryotic neomycin-resistance-encoding gene acts as a transcriptional silencer in eukaryotic cells. *Gene*, **99**, 249–254.

Ausubel, F.N., Brent, R., Kingston, R.E., Moore, D.D., Seichmann, J.G., Smith, J.A. and Struhl, K. (1987) *Current Protocols in Molecular Biology*. New York: John Wiley and Sons.

Bäumlein, H., Wobus, U., Pustell, J. and Kafatos, F.C. (1986) The legumin gene family: structure of a B-type gene of *Vicia faba* and a possible legumin gene specific regulatory element. *Nucl. Acids Res.* **14**, 2707–2720.

Bäumlein, H., Boerjan, W., Nagy, I., Panitz, R., Inzé, D. and W bus, U. (1991a) Upstream sequences regulating legumin gen expression in heterologous transgenic plants. *Mol. Gen. Genet.* **225**, 121–128.

Bäumlein, H., Boerjan, W., Nagy, I., Bassüner, R., Van Montagu, M., Inzé, D. and W bus, U. (1991b) A novel seed protein gene

from *Vicia faba* is developmentally regulated in transgenic tobacco and *Arabidopsis* plants. *Mol. Gen. Genet.* **225**, 459–467.

Bust s, M.M., Guiltinau, M.J., J rdano, J., Begum, D., Kalkan, F.A. and Hall, T.C. (1989) Regulation of β-glucuronidase expression in transgenic tobacco plants by an A/T rich *cis*-acting sequence found upstream of a french bean β-phaseolin gene. *Plant Cell*, **1**, 839–853.

Chen, Z.L., Naito, S., Nakamura, I. and Beachy, R.N. (1988) Regulated expression of genes encoding soybean β-conglycinin in transgenic plants. *Devel. Genet.* **10**, 112–122.

Dickinson, C.D., Evans, R.P. and Nielsen, N.C. (1988) RY repeats are conserved in the 5′-flanking region of legume seed protein genes. *Nucl. Acids Res.* **16**, 371.

Ebert, P.R., Ha, S.B. and An, G. (1987) Identification of an upstream element in the nopaline synthase promoter by stable and transient assays. *Proc. Natl Acad. Sci. USA*, **84**, 5745–5749.

G ldberg, R.B., Barker, S.J. and Perez-Grau, L. (1989) Regulation of gene expression during plant embryogenesis. *Cell*, **56**, 149–160.

Harada, J.J., Barker, S.J. and Goldberg, R.B. (1989) Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell*, **1**, 415–425.

Herman, L.M.F., Van Montagu, M. and Depicker, A.G. (1986) Isolation of tobacco DNA segments with plant promoter activity. *Mol. Cell. Biol.* **6**, 4486–4492.

J fferson, R., Goldbrough, A. and Bevan, M. (1990) Transcriptional regulation of a patatin-1 gene in tobacco. *Plant Mol. Biol.* **14**, 995–1006.

J fuku, K.D., Okamuro, J.K. and Goldberg, R.B. (1987) Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. *Nature*, **328**, 734–737.

Jordano, J., Almoguera, C. and Thomas, T.L. (1989) A sunflower helianthinin gene upstream sequence ensemble contains an enhancer and sites of nuclear protein interaction. *Plant Cell*, **1**, 855–866.

K ncz, C., Martini, N., Mayerhofer, R., Koncz-Kalman, Z., Körber, H., Redei, G.P. and Schell, J. (1989) High-frequency T-DNA-mediated gene tagging in plants. *Proc. Natl Acad. Sci.* **86**, 8467–8471.

McCarty, D.R. and Carson, C.B. (1991) The molecular genetics of seed maturation in maize. *Physiol. Plant.* **81**, 267–272.

Memelink, J., de Pater, B.S., Hoge, J.H.C. and Schilperoort, R.A. (1987) T-DNA hormone biosynthetic genes: phytohormones and gene expression in plants. *Devel. Genet.* **8**, 321–337.

Müller, M.M., Gerster, T. and Schaffner, W. (1988) Enhancer sequences and the regulation of gene transcription. *Eur. J. Biochem.* **176**, 485–495.

Mundy, J., Yamaguchi-Shinozaki, K. and Chua, N.H. (1990) Nuclear proteins bind conserved elements in the abscisic acid responsive promoter of a rice rab gene. *Proc. Natl Acad. Sci. USA*, **87**, 1406–1410.

Okamuro, J.K. and Goldberg, R.B. (1989) Regulation of gene expression: general principles. In *The Biochemistry of Plants*, Volume 15. San Diego: Academic Press, pp. 1–82.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A. and Saedler, H. (1987) The regulatory *C1* locus of *Zea mays* encodes a protein with homology to *myb* proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J.* **6**, 3553–3558.

Platt, S.G. and Yang, N.S. (1987) Dot assay for neomycin phosphotransferase activity in crude cell extracts. *Anal. Biochem.* **162**, 529–535.

Reiss, B., Sprengel, R., Will, H. and Schaller, H. (1984) A new sensitive method for quantitative and qualitative assay of neomycin phosphotransferase in crude cell extracts. *Gene*, **30**, 211–218.

Rerie, W.G. (1989) *The structure of pea legumin genes and their expression in transgenic tobacco.* Ph.D. Thesis. Canberra: Australian National University.

Rerie, W.G., Whitecross, M. and Higgins, T.J.V. (1991) Developmental and environmental regulation of pea legumin genes in transgenic tobacco. *Mol. Gen. Genet.* **225**, 148–157.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning – a Laboratory Manual.* New York: Cold Spring Harbour Laboratory Press.

Schubert, R., Bäumlein, H., Czihal, A. and Wobus, U. (1990) Genomic sequence of a 12 S seed storage protein gene from oat (*Avena sativa* L. cv. 'Solidor'). *Nucl. Acids Res.* **18**, 377.

Shirsat, A.H., Wilford, N., Croy, R.R.D. and Boulter, D. (1989) Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. *Mol. Gen. Genet.* **15**, 326–331.

Thompson, A.J., Bown, D., Yaish, S. and Gatehouse, J. (1991) Differential expression of seed storage protein genes in the pea legJ subfamily, sequence of the gene legK. *Biochem. Physiol. Pflanzen*, **187**, 1–12.

Weising, K. and Kahl, G. (1991) Towards an understanding of plant gene regulation: the action of nuclear factors. *Z. Naturforsch.* **46c**, 1–11.

Wobus, U., Bäumlein, H., Inzé, D. and Nagy, I. (1989) *Vicia faba* storage protein genes and their promoter activity in transgenic tobacco plants. In *Applied Plant Molecular Biology* (Galling, G., ed.). Braunschweig: Technical University, pp. 98–103.

Zenke, M., Grundström, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. and Chambon, P. (1986) Multiple sequence motifs are involved in SV40 enhancer function. *EMBO J.* **5**, 387–397.

EMBL Data Library accession number X03677.

MGG

Molecular & General Genetics

An International Journal

Continuation of Zeitschrift für Vererbungslehre The First Journal on Genetics Founded in 1908

Volume 225 1991

Springer International

Mol Gen Genet (1991) 225:459–467
002689259100061S

MGG

A novel seed protein gene from *Vicia faba* is developmentally regulated in transgenic tobacco and *Arabidopsis* plants

Helmut Bäumlein[1], Wout Boerjan[2], Istvan Nagy[1,3], Ronald Bassüner[1], Marc Van Montagu[2], Dirk Inzé[2] and Ulrich Wobus[1]

[1] Zentralinstitut für Genetik und Kulturpflanzenforschung der AdW, O-4325 Gatersleben, FRG
[2] Laboratorium voor Genetica, Rijksuniversiteit Gent, B-9000 Gent, Belgium
[3] Agricultural Biotechnology Center, H-2101 Gödöllö, Hungary

Received September 9, 1990

Summary. We have isolated a novel gene, denoted *USP*, from *Vicia faba* var. minor, which corresponds to the most abundant mRNA present in cotyledons during early seed development; however, the corresponding protein does not accumulate in cotyledons. The characterized *USP* gene with its two introns is 1 of about 15 members of a gene family. A fragment comprising 637 bp of 5′ flanking sequence and the total 5′ untranslated region was shown to be sufficient to drive the mainly seed-specific expression of two reporter genes, coding for neomycin phosphotransferase II and β-glucuronidase, in transgenic *Arabidopsis thaliana* and *Nicotiana tabacum* plants. We showed that the *USP* promoter becomes active in transgenic tobacco seeds in both the embryo and the endosperm, whereas its activity in *Arabidopsis* is detectable only in the embryo. Moreover, we demonstrated a transient activity pattern of the *USP* promoter in root tips of both transgenic host species.

Key words: *Arabidopsis thaliana* – β-Glucuronidase – Root tip – Seed protein gene – *Vicia faba*

Introduction

Despite their apparent morphological simplicity, plants express organ-specific and developmentally regulated genetic programs comparable in complexity to those in animal systems (Goldberg 1988; Goldberg et al. 1989). In an attempt to understand this complexity, the gene families coding for seed proteins have been used as a model experimental system in plant molecular biology (Goldberg et al. 1989). These studies are elucidating the basic principles of the regulation of gene expression during embryogenesis and may provide information that can be applied to improvement of seed protein quality.

As in most dicotyledonous plants, the seed storage protein fraction of the fava bean *Vicia faba* var. minor is dominated by its globulin components (Müntz et al. 1986). Structural and functional data are available for the gene families coding for the major 11S legumin and the 7S vicilin proteins (Wobus et al. 1986; Bäumlein et al. 1986, 1987; Weschke et al. 1987; Heim et al. 1989).

We have recently described cDNA clones specific for an Unknown Seed Protein (*USP*) (Bassüner et al. 1988). The corresponding genes are transcribed into the most abundant mRNA present during early seed development with a time profile similar to that of vicilin mRNA. In spite of the abundance of the mRNA, a similarly abundant protein of the expected size (30 kDa) has not been found (Bassüner et al. 1988). This observation indicates that expression levels are controlled both by transcriptional and extensive post-transcriptional processes. As a first step in revealing the underlying regulatory mechanisms we sequenced a *USP* gene and its flanking regions. In addition, we fused the *USP* promoter region to bacterial reporter genes, and describe the complex seed and root tip-specific expression in transgenic tobacco and *Arabidopsis* plants revealed by the histochemical colour reaction for β-glucuronidase (GUS) activity.

Materials and methods

Gene isolation and sequencing. The recombinant phage Vf30.1 was originally isolated from a phage library of the field bean (*V. faba* var. minor) genome (Bäumlein et al. 1986), using a *USP*-specific cDNA as probe (Wobus et al. 1986). A 3.5 kb *Pst*I fragment containing a member of the USP gene family was subcloned in the phage vector M13mp18 for sequencing and in pUC18 for use in further constructions.

The M13 phage insert was sequenced in both orientations using systematic deletions (Hong 1982) and the chain termination method (Sanger et al. 1977). For the processing of the sequence data a modified Pustell computer program was used (Pustell and Kafatos 1986). The transcription start site was determined by primer extension (Ausubel et al. 1987) using the synthetic primer

Offprint requests to: M. Van Montagu

5′CAAACTCCATTTGACTGGCT3′. A known sequence ladder was used as size marker (Fig. 2).

Plasmid constructs and generation of transgenic plants. For the construction of chimaeric genes consisting of the 5′ sequences (−637 to +51) of the *USP* gene fused to reporter genes encoding neomycin phosphotransferase II (NPTII) or GUS, a unique *Bst*XI site at the ATG start codon was converted into a *Bgl*II site by the insertion of a *Bgl*II linker into the blunt-ended *Bst*XI site and the coding region together with the 3′-flanking sequence was removed. From the resulting plasmid, the flanking and the 5′-nontranslated regions can be obtained as a 680 bp *Bam*HI-*Bgl*II fragment. This fragment was cloned in the appropriate orientation upstream of the promoterless *nptII* gene in the intermediate vector pGV180. This vector is a derivative of pGV150 conferring hygromycin resistance (Herman et al. 1986). For the *uid*A construct the *Bam*HI-*Bgl*II fragment was blunt-ended and cloned into the blunt-ended *Sal*I site of pGUS1 (Peleman et al. 1989). The resulting *Hin*dIII-*Bam*HI fragment containing the chimaeric *USP-uid*A fusion was used to replace the *Hin*dIII-*Bgl*II fragment of the binary vector pGA472 (An et al. 1985). The plasmids were transferred into the *Agrobacterium* strain pGV2260 (Deblaere et al. 1985) by triparental mating and used for the transformation of *Nicotiana tabacum* cv. Havana by the leaf disc method (Horsch et al. 1985) and *Arabidopsis* ecotype Columbia by the root transformation method (Valvekens et al. 1988). The integrity of all constructs was checked both in *Agrobacterium* and in plants by Southern hybridization (data not shown).

Genomic blots. To determine the size of the gene family, 10 µg genomic *V. faba* DNA was digested with an excess of *Eco*RI, *Bam*HI, *Hin*dIII, *Bgl*II, or *Sph*I. Lambda DNA added to an aliquot of the reaction mixture was used to check that digestion was complete. After blotting on CCA paper (Hunger et al. 1986) the filter was hybridized (2×SSC, 65° C) with the 3.5 kb *Pst*I fragment labelled to a specific activity of 10^8 cpm/µg by random priming (Feinberg and Vogelstein 1983) and exposed to X-ray film for 2 days.

NPTII and GUS assays. NPTII activity was determined from 100 mg tissue. Equal amounts of protein (determined according to Bradford 1976) from each extract, were assayed by gel electrophoresis (Reiss et al. 1984). For the analysis of the tissue specificity of the promoter, tobacco seeds were hand dissected into embryo and endosperm.

GUS assays were performed basically as described by Jefferson (1987).

For the histochemical analysis, mature seeds were imbibed for 4 h and embedded in a 5% agarose solution in water without fixation. The block of agarose containing the seed was cut in the desired orientation with a scalpel and fixed with Pattex Super Gel (S.A. Henkel N.V., Belgium) on the mounting table of a Vibroslicer (Laborimpex, Belgium). After slicing, the embryo and endosperm in the sections were separated. The sections (approximately 30 µm) were each placed in a drop of 50 mM phosphate buffer, pH 7.0 containing 1 mM 5-bromo-4-chloro-3-indolyl-β-D-glucuronide (X-gluc), 0.1 mM potassium ferricyanide, and 0.1 mM potassium ferrocyanide in a petri dish. Sections were incubated at 37° C for 10 min to 24 h in a humidified chamber and then mounted on microscope slides for photography. Photographs were taken with a Wild MPS51 microscope (Heerbrugg, Switzerland).

Gene expression studies in intact plants were done using sterile seedlings (grown in a 16 h light/8 h dark cycle on K1 medium; Valvekens et al. 1988) that had been placed directly in X-gluc solution or had been cut at leaf, cotyledon, hypocotyl, and root prior to incubation to avoid penetration problems. Embryos were dissected by hand from mature seeds 4 h after imbibition and incubated in X-gluc at 37° C.

Results

Structure of the USP *gene*

Screening of a *V. faba* var. minor genomic library with a *USP* cDNA clone revealed two positive phages one of which (λUSP30.1) was chosen for restriction mapping and sequence analysis. The nucleotide sequence of the 3.5 kb *Pst*I fragment is shown in Fig. 1. By comparison with cDNA sequences (Bassüner et al. 1988), two introns, 81 and 110 bp in length, were localized. The border sequences of both introns obey the consensus rules derived for plant genes (Brown 1986). A comparison with cDNA clones pUSP87 and pVfc13 (Bassüner et al. 1988) locates the polyadenylation site 251 bp downstream from the TAA stop codon (Fig. 1). This 3′-untranslated region contains multiple and overlapping polyadenylation signals with plant-specific features (Joshi 1987a) 10 to 31 bp upstream from the poly(A) site. The TGTGTTT motif often found in 3′-flanking regions of plant genes (Joshi 1987a) immediately precedes the poly(A) site.

The transcription start site of the *USP* gene was determined by primer extension experiments (Figs. 1 and 2). As shown in Fig. 2, two bands of equal intensity mark either an A or a C as the potential cap site. Since 67 out of 79 plant genes listed by Joshi (1987b) employ A at the transcription initiation site, we chose the A as position +1 of the *USP* gene. Conceptual translation of the mRNA defines a polypeptide with no obvious overall homology to any other protein sequence present in protein databases (release 21 of the PIR and release 12 of SWISS-PROT). We note however that the signal sequence coding region (+52 to +198; see Bassüner et al. 1988) is interrupted by the first intron as in a tomato proteinase inhibitor I gene (Lee et al. 1986) and that in both genes sequences up to the intron are remarkably homologous: 28 out of 39 nucleotides are identical and 8 out of 13 encoded amino acid residues are functionally equivalent (a two-codon "deletion" in the *USP* gene is not counted; data not shown).

A search of the 5′ upstream region revealed, besides a TATA box at approximately 30 bp upstream from

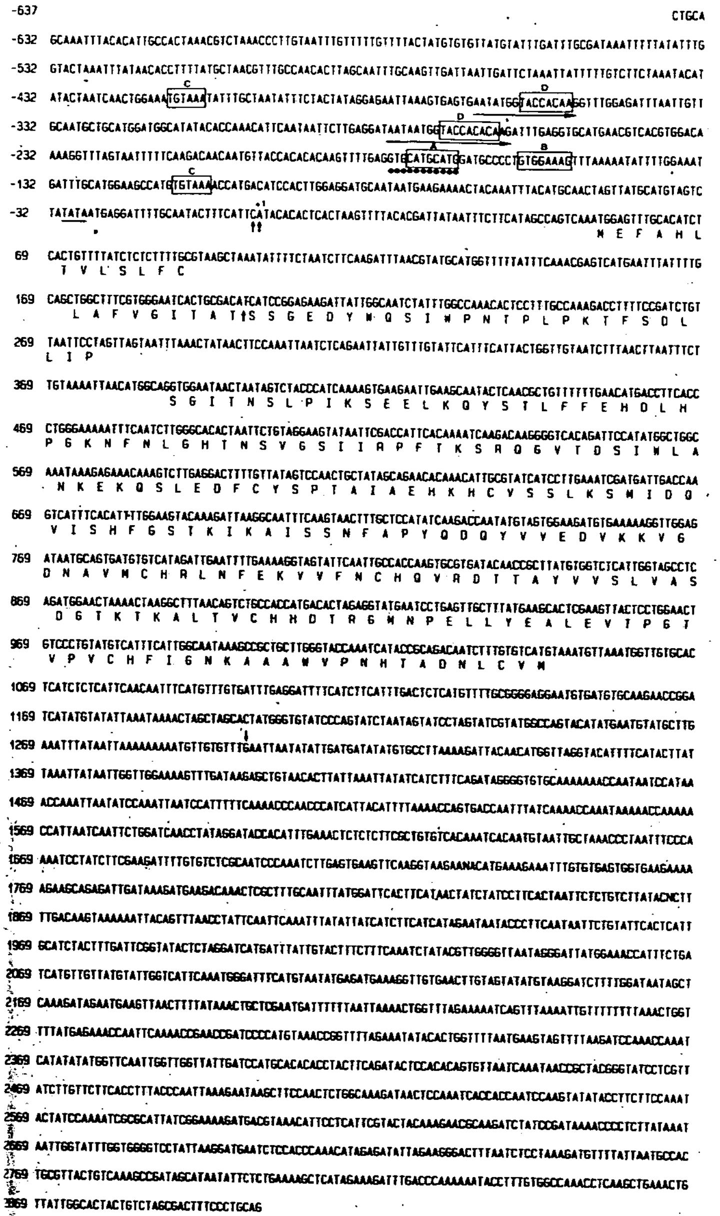


Fig. 1. Sequence of the *USP* gene. The coding region is interrupted by two introns. The *vertical arrow* in the protein sequence marks the position of the presumed signal peptide cleavage point; the *vertical arrow* in the 3′ end of the gene indicates the polyadenylation site, and the *two vertical arrows* in the 5′ region mark the transcription start sites. The CATGCATG box, (A) is part of an 11 bp purine-pyrimidine repeat represented by *dots*. Box B, SV40 core enhancer motif; box C, prolamin box; box D, a sequence present in front of several soybean seed protein genes, which is part of an imperfect direct repeat. The putative TATA box is *underlined* and is located approximately 30 bp upstream from the transcription start

the transcription initiation site, a number of conserved sequence elements (Fig. 1): the seed protein gene-specific CATGCATG motif, which is part of a purine-pyrimidine repeat (Dickinson et al. 1988) at positions −170 to −177, the SV40 core enhancer motif (−154 to −160)

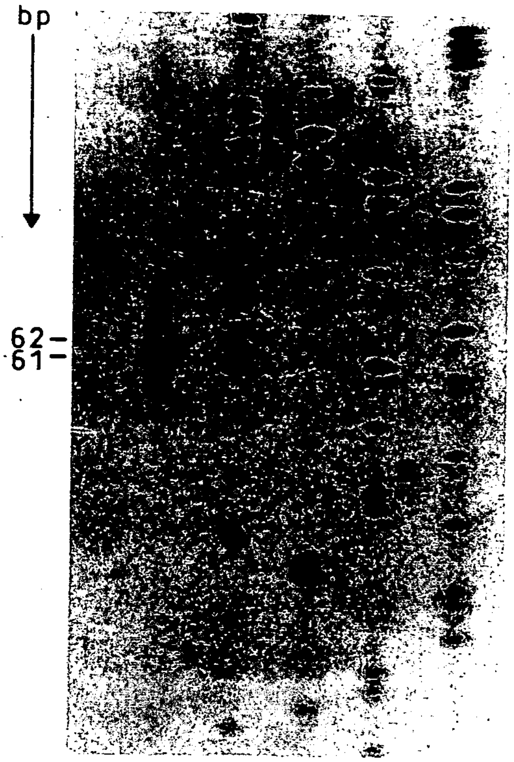


Fig. 2. Determination of the transcription start site by primer extension. Using a known sequencing ladder as marker, the size of the primer extension products were 61 and 62 nucleotides long (see Materials and methods)

GTGGAAAG (Benoist and Chambon 1981), and two copies of the so-called prolamin box (Colot et al. 1987; Matzke et al. 1990) at positions −108 to −113 and −409 to −414. The sequence motifs TACCACAA and TACCACACA found as parts of an imperfect direct repeat in the *USP* promoter (positions −262 to −272 and −352 to −360) closely resemble another sequence element found upstream of several soybean seed protein genes (Goldberg 1986). The importance of all these structurally conserved short sequences for the tissue-specific and development-dependent activity of the *USP* promoter remains to be determined.

The USP *gene family*

Sequence analysis of *USP* cDNA clones had already proven the existence of several *USP* genes (Bassüner et al. 1988). Therefore, genomic hybridization experiments were carried out to estimate roughly the total number of *USP* genes in the *V. faba* var. minor genome, as shown in Fig. 3. By using different restriction enzymes that do not cut within the sequenced *USP* gene we detected approximately 15 hybridizing bands, suggesting that the *USP* gene family consists of roughly 10 to 20 members. This is about the number determined for the legumin *B* gene subfamily (Heim et al. 1989).

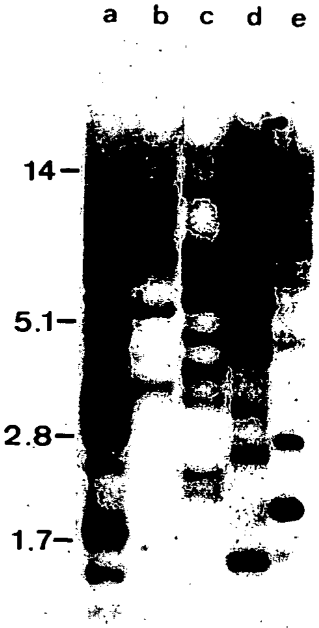


Fig. 3. Genomic blot of *Vicia faba* DNA digested with different restriction enzymes and hybridized to a fragment containing the *USP* gene. a, *Eco*RI; b, *Bam*HI; c, *Hin*dIII; d, *Bgl*II; e, *Sph*I. The size markers shown on the left are in kb

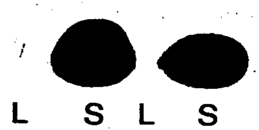


Fig. 4. Comparison of NPTII activity in equal amounts of protein from extracts of seeds (S) and leaves (L) in two independent tobacco transformants

The USP *5′-flanking region confers seed-specific expression on two reporter genes*

We fused a *Bam*HI-*Bgl*II fragment containing 637 bp of the 5′-flanking and the total 5′-untranslated region of 51 bp to a promoterless *nptII* gene in the intermediate vector pGV180 (see Materials and methods) and used it to transform *N. tabacum*. All ten hygromycin-resistant tobacco plants regenerated carried one to five copies of the chimaeric gene (data not shown). With few quantitative differences, all plants produced high levels of NPTII only in seeds as shown in Fig. 4. We never found NPTII activity in leaves, even with a tenfold higher amount of total protein in the assay. Plants transformed with pGV180 alone were NPTII negative in both seeds and leaves. In order to determine the distribution of promoter activity in seed tissues, we hand-dissected developing seeds transformed with the *USP-nptII* fusion construct into embryo and endosperm. Microscopic examination showed contamination of some of the embryos by endosperm but this did not exceed 10%. On the basis of equal protein concentrations we found approximately three- to fourfold higher NPTII activity in embryos as compared with endosperm (mean of three experiments, data not shown). These results were confirmed by analysing hand-prepared embryos and endosperm of seeds transformed with a completely different vector (pGA472) containing the *uid*A gene (coding for β-glucuronidase; Jefferson 1987) fused to the *USP* promoter fragment mentioned above (see Materials and methods). The mean GUS activity in two experiments with samples from five different plants was tenfold higher in embryos than in endosperm. In spite of the quantitative differences between the two experimental series, it is evident that the *USP* promoter is not only active in the embryo but also in the endosperm of tobacco.

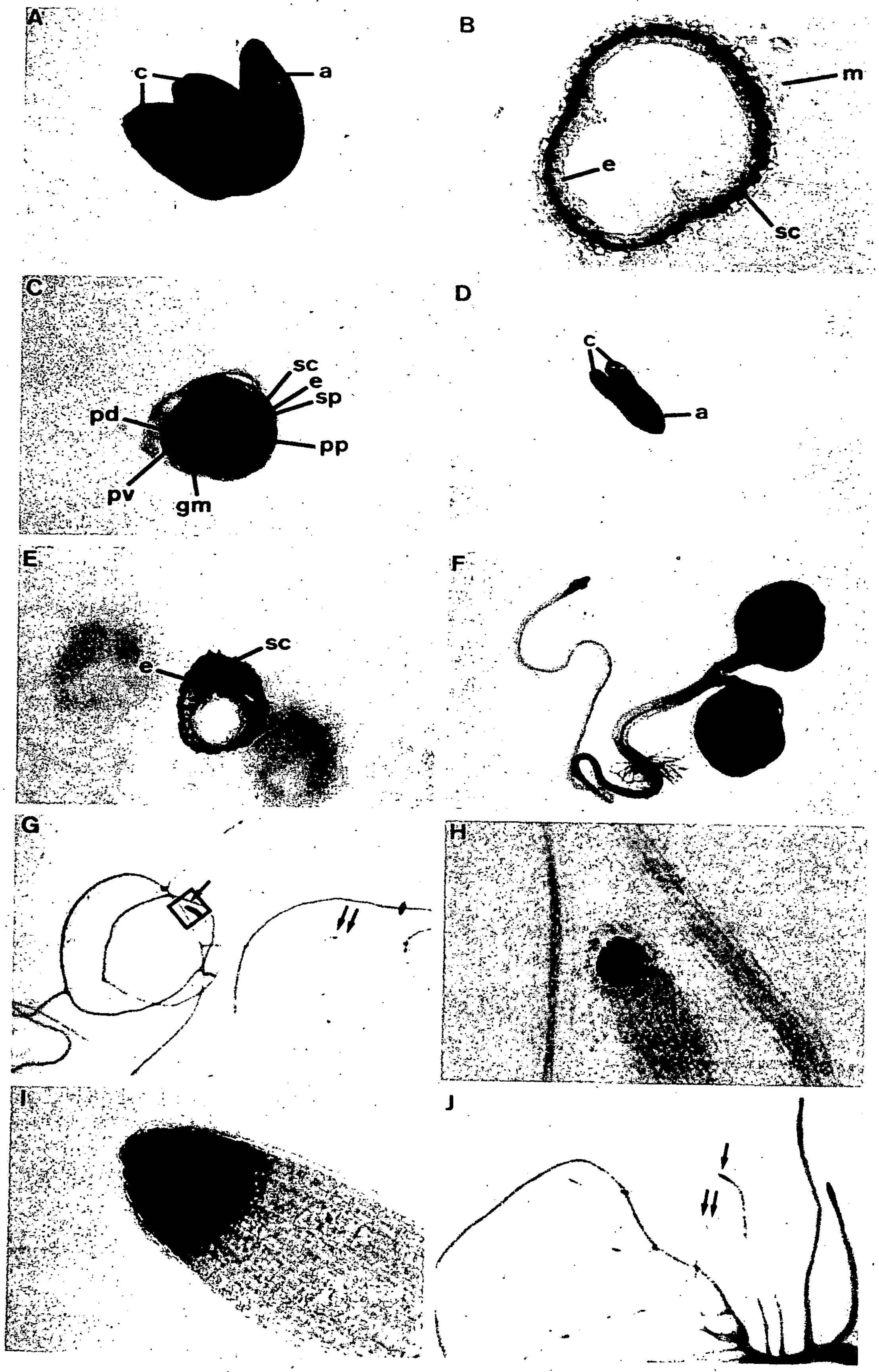
Histochemical localisation of GUS activity in seeds and seedlings of transgenic Arabidopsis *and tobacco plants*

Arabidopsis thaliana ecotype Columbia and *N. tabacum* cv. Havana were transformed with a *USP-uid*A fusion gene (see above and Materials and methods) and F_1 seeds and seedlings were analysed for GUS activity. Initially, the F_1 progeny of different independent transgenic lines were analysed. Since the resulting overall staining patterns were similar (data not shown), detailed histochemical analysis was focused on the progeny of two different transgenic lines from each of the two host plants.

GUS activity in seeds. There is a substantial difference in the ability of the GUS substrate (X-gluc) to penetrate different tissues. Seeds especially do not take up X-gluc easily. Therefore, it is difficult to obtain reliable results by staining intact tissues. The problem can be partly circumvented by using a Vibroslicer (see Materials and methods). With this device, unfixed mature seeds can be cut into thin sections prior to the histochemical reaction.

When sections of transformed *Arabidopsis* seeds were incubated in X-gluc, all cells of the embryo as well as the cell layer between the testa and the embryo stained blue. This layer is the outermost cell layer of the endosperm, the so-called aleurone layer. The rest of the endo-

Fig. 5A–J. Histochemical localization of GUS activity in embryo, endosperm and seedlings of *Arabidopsis* and *Nicotiana*. **A** Hand-prepared embryo of *Arabidopsis*. **B** Seed coat and endosperm of *Arabidopsis*; section through mature seed. **C** Section through mature seed of *Arabidopsis*. **D** Hand-prepared embryo of *Nicotiana*. **E** Seed coat and endosperm of *Nicotiana*; section through mature seed. **F** Six-day-old seedling of *Arabidopsis*. **G** Part of root system of *Arabidopsis*; the root with blue root tip is itself a lateral root (not visible on picture). **H** Enlargement of *box* in **G**, showing GUS-positive root tip of *Arabidopsis*. **I** Distal part of a side root of *Nicotiana*. **J** Root system of *Nicotiana* (20 days old). a, axis; c, cotyledons; e, endosperm; gm, ground meristem; m, mucilage; pd, protoderm; pp, palisade parenchyma; pv, provascular bundle; sc, seed coat; sp, spongy parenchyma. *Arrow*, GUS-positive root tip; *double arrow*, GUS-negative root tip


A
c
a
B
m
e
sc
C
sc
e
sp
pd
pp
pv
gm
D
c
a
E
sc
e
F
G
H
I
J

sperm tissue is resorbed by the developing embryo as the seed matures (Vaughan et al. 1971; Bouman 1975).

However, when embryo and endosperm were separated after slicing, but before incubation, only embryo cells showed GUS activity (Fig. 5A, B). The initial result using intact tissue slices can be explained by diffusion of the primary reaction product from cells where it is produced to cells where it precipitates (see Jefferson 1987). However, when an oxidative catalyst is used (see Materials and methods), precipitation of the product appears much faster and more localized. In the presence of the oxidative catalyst, sections with embryo and endosperm do not stain in the endosperm. The reaction is seen first in the palisade parenchyma cells of the cotyledons. Subsequently, blue precipitate is localized in the provascular bundle and protoderm of the axis. After prolonged incubation, weaker blue stain is present in all other cells of the embryo (Fig. 5C).

In contrast to *Arabidopsis* seeds, when tobacco seed slices were stained after separation of endosperm and embryo, GUS activity could be detected in both tissues, thus demonstrating a difference in tissue-specific expression of the chimaeric *USP-uid*A gene in the two heterologous hosts *Arabidopsis* and tobacco (Fig. 5D, E).

GUS activity in seedlings. To analyse the distribution of GUS activity during early plant development, seedlings were taken at different times after germination and stained in X-gluc solution.

Germinating seedlings of *Arabidopsis* turn completely blue upon overnight incubation with X-gluc. Shorter incubation periods reveal differences. Whereas the cotyledons show patches of blue stain, the radicle is uniformly dark blue (data not shown). In germinated seedlings (fully expanded cotyledons), the blue precipitate can be detected in the root with decreasing intensity towards the root tip, whereas the root tip itself is deep blue. Root hairs also contain the precipitate. Upon longer incubation, blue precipitate can also be observed in the vascular bundle, mesophyll cells and epidermal cells of the cotyledons and the hypocotyl (Fig. 5F).

At a later developmental stage, when the first true leaves occur, GUS activity is still present in the root, but the staining is much weaker. It is not detected in leaves, even when they are cut to facilitate penetration of the substrate. In 10% of the seedlings, weak blue staining is also detected in the vascular bundle of the root. The root cap cells and the root meristem show the highest GUS activity.

During the early stages of secondary root formation (up to approximately ten roots per individual plant), GUS activity can only be detected in the cotyledons and in a low percentage of the root tips. We observed that at this stage 10%–20% of the plants no longer show root tip activity, 70%–80% of the plants show GUS activity in only a few root tips (ranging from 10% to 90% of the root tips), and in 10%–20% of the plants, all of the root tips turn blue (Fig. 5G, H). At later stages in development, up to bolting, these relative numbers shift towards plants having no root tip activity at all. GUS activity in the root tips does not seem to be correlated with the age of the side roots, since young side roots do not always have this activity. Roots of mature plants were always GUS negative. In all experiments described, the root system was cut just below the hypocotyl prior to incubation to avoid contamination due to GUS activity in other parts of the plant, e.g. the hypocotyl or cotyledons.

In *N. tabacum*, the pattern of GUS activity in seedlings is very similar to that of *Arabidopsis*. The root tip activity is initially very strong (Fig. 5I). When the plant matures (two-leaf stage), activity is detected mainly in the root cap cells but only after prolonged incubation in X-gluc solution (30 h), suggesting a decrease in the activity of the *USP-uid*A construct. As described for *Arabidopsis*, this activity is not always present in all root tips. The individual plants can have root tips that stain blue, as well as root tips that remain uncoloured (Fig. 5J).

Discussion

Embryo versus endosperm activity of the USP *promoter*

To analyse the tissue-specific promoter activity within seeds we transformed both *Arabidopsis* and tobacco plants with two constructs containing either the *nptII* gene or the *uid*A gene driven by the *USP* promoter. The constructs placed in different vector plasmids share only the *USP* promoter fragment, thereby ruling out possible artificial influences of the neighbouring T-DNA sequences on the expression of the chimaeric genes. When isolated embryo and endosperm tissues from tobacco seeds were tested for enzyme activity, both tissues were found to give positive results with both constructs but with appreciably less activity in the endosperm. The GUS analyses were further confirmed at the histochemical level. In contrast to these results, *Arabidopsis* seeds transformed with the *USP-uid*A construct show activity only in the embryo but not in the endosperm cell layer. Within the embryos of both species all cells turn completely blue upon incubation with X-gluc if no catalyst is used. Addition of ferro-ferricyanide as an oxidative catalyst remarkably enhances the cell specificity of the reaction. First, a reaction is seen in the palisade parenchyma cells of the cotyledons and in the provascular bundle and protoderm of the axis. After longer incubation, weak blue staining is localized in all other cells of the embryo. This weak blue staining could reflect weaker expression of the *USP-uid*A gene in these cells, but could also be due to diffusion of the breakdown product of X-gluc. Similar experiments were carried out with seeds of *Arabidopsis* transformed with a chimaeric gene consisting of the regulatory sequences of *at2S-1*, one of the four genes coding for the 2S napins in *Arabidopsis* (Krebbers et al. 1988a), coupled to the coding sequence of the *uid*A gene. In these thin sections the reaction was first detected in the endosperm. Subsequently, blue precipitate was observed in the spongy parenchyma cells of the cotyledons and the ground meristem cells of the axis (W. Boerjan, unpublished results).

These results indicate that the spatial expression pattern observed with the *USP* promoter is very specific and reflects the amount of β-glucuronidase present in the different cell types, and is not due to artefacts such as penetration and differences in cell size.

The difference in the expression pattern of a heterologous gene in two plant species may reflect differences in regulation due to the ability/inability of *trans*-acting factors to recognize the *cis*-acting elements of the foreign *USP* promoter in the respective endosperm. In any case, data obtained from gene expression studies in heterologous host plants should be interpreted with caution and cannot be generalized. The situation is further complicated by the fact that we do not know how the *USP* promoter behaves in the homologous *V. faba* background, because there is only a rudimentary endosperm in legumes, which disappears at early developmental stages. The only available data pertinent to the problem are for soybean. In this species two seed protein mRNAs coding for Kunitz trypsin inhibitor and β-conglycinin cannot be detected in the endosperm before it disappears during embryogenesis (Perez-Grau and Goldberg 1989). This observation suggests that seed storage protein genes of legumes become active exclusively in the embryo and, as several studies have shown (for a review see Goldberg et al. 1989), that this behaviour is maintained in transgenic tobacco plants. However, this is not true for all investigated genes, since legumin genes *LegA* from pea (Croy et al. 1988) and *LeB4* from fava bean (Wobus et al. 1989; Bäumlein et al. 1990) are active in both the tobacco embryo and endosperm. Such differences in tissue-specific expression between different seed protein genes are most obviously demonstrated in the case of two β-conglycinin genes of soybean: one is active only in the tobacco embryo (Barker et al. 1988), another in both the embryo and the endosperm (S.J. Barker and R.B. Goldberg, unpublished results).

Histochemical localisation of GUS activity in seedlings of transgenic Arabidopsis *and tobacco plants*

There is a substantial difference in the ability of the GUS substrate to penetrate different tissues. Roots, for example, take up the substrate very fast in comparison with the cotyledons, which in fact need to be cut or squashed to enable adequate penetration. Therefore it is difficult to use histochemical techniques to obtain an idea about the absolute levels of GUS enzyme, let alone promoter strength, when comparing different tissues. Another parameter that impedes attempts to correlate directly the appearance of the blue precipitate with promoter activity is the high stability of the β-glucuronidase. As the half-life of GUS in germinating seeds is about 48 h (Bustos et al. 1989) it is difficult to determine whether the blue colour detected in seedlings is due to the presence of stable enzyme synthesized during seed development or to newly formed GUS. We tend to attribute the GUS activity in cotyledons and roots (with the exception of the activity in the root tip; see below) of young seedlings to stable enzyme synthesized previously during embryogenesis, but cannot exclude reactivation of the promoter.

Variation in root tip activity

In transformed *Arabidopsis* as well as tobacco plants, root tip activity is strong in the early stages of development (1- to 2-week-old seedlings) and ceases upon maturation. Moreover, plants that already have a well-developed root system with secondary roots do not always show activity in all root tips. Roots showing root tip activity are morphologically indistinguishable from roots lacking this activity.

The fact that side roots can also have root tip activity suggests that this activity is due to de novo synthesis and not to a residual or redistributed activity of β-glucuronidase synthesized during embryogenesis.

There are several possible explanations for the root tip activity of the *USP* promoter. First, this activity may indicate that the *USP* gene has a specific, but unknown function in root tip tissues. Second, the transient state of activity might well be considered as an evolutionary relic without detrimental effect, but also without functional value. Third, the root tip activity of the *USP* promoter might be due to the structure of the chimaeric gene: it is possible that not all of the control elements needed for *USP* gene expression reside in the 680 bp 5'-flanking region. Fourth, the transfer of the chimaeric gene to a heterologous host, which possibly lacks silencing activities, could also be the cause of the unexpected expression. This explanation is less likely, because experiments with *Arabidopsis* plants transformed with a chimaeric *uid*A gene consisting of the 5'-regulatory sequences of the *ats-1* gene (one of the four genes coding for the small subunit of ribulose-1,5-bisphosphate carboxylase in *Arabidopsis*; Krebbers et al. 1988b), also show root tip activity (W. Boerjan, unpublished results). A similar result has been obtained with transgenic *Arabidopsis* plants that express a chimaeric *at2S-1-gus* gene (see above) (W. Boerjan, unpublished results), thus showing that the expression of certain chimaeric genes in root tips might be a common phenomenon and not due merely to the transfer from one species to another. To exclude the possibility that the root tip activity is due to the structure and/or the transfer of the chimaeric gene, in situ hybridization experiments need to be carried out to determine whether the *USP* mRNA is present in the root tips of *V. faba* itself.

A second question is why not all root tips of a well-developed root system show GUS activity. One might speculate that the root tips are in a developmentally or physiologically different state. The variation in root tip activity could reflect a subtle balance in the interaction between regulatory factors, which may be distorted in a fast-dividing tissue, for example as a result of titration of regulatory factors. Alternatively, one might argue that position effects and copy number differences between the seeds of one transformant with multiple copies play a role in the variation. This explanation is unlikely, because F_1 progeny of a transformed *Arabidopsis* plant,

containing one copy of the *ats-1* gene, which was made homozygous for the T-DNA, also show variation in root tip activity (W. Boerjan, unpublished results).

Currently, we are analysing progressive promoter deletions to identify the *cis* elements responsible for the seed- and root tip-specific expression of the *USP* gene.

Acknowledgements. We would like to thank Andreas Czihal, Elke Liemann, Thea Tölle, Hannelore Weiß, Lieve Ongena, Marianne Habets, and Angela Stegmann for excellent technical assistance. H.B. would like to thank Prof. Kafatos, in whose laboratory the *USP* gene was sequenced, and Drs. Allan Caplan, Diane Jofuku, Jack Okamuro, and Gilbert Engler for helpful discussions. M. De Cock, K. Spruyt, S. Van Gijsegem, and V. Vermaercke are thanked for preparing the manuscript and illustrations. This research was supported by grants of the Services of the Prime Minister (I.U.A.P. 12OC0187), the ASLK-Kankerfonds, and the Kombinat Pflanzenzüchtung und Saatgutwirtschaft Quedlinburg. D.I. is a Senior Research Assistant of the National Fund for Scientific Research (Belgium) and W.B. is indebted to the I.W.O.N.L. for a predoctoral fellowship.

References

An G, Watson RD, Stachel S, Gordon MP, Nester EW (1985) New cloning vehicles for transformation of higher plants. EMBO J 4:277–284

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1987) Current protocols in molecular biology. Wiley Interscience, New York

Barker SJ, Harada JJ, Goldberg RB (1988) Cellular localization of soybean storage protein mRNA in transformed tobacco seeds. Proc Natl Acad Sci USA 85:458–462

Bassüner R, Bäumlein H, Huth A, Jung R, Wobus U, Rapoport TA, Saalbach G, Müntz K (1988) Abundant embryonic mRNA in field bean (*Vicia faba* L.) codes for a new class of seed proteins: cDNA cloning and characterization of the primary translation product. Plant Mol Biol 11:321–334

Bäumlein H, Wobus U, Pustell J, Kafatos FC (1986) The legumin gene family structure of a B type gene of *Vicia faba* and a possible legumin gene specific regulatory element. Nucleic Acids Res 14:2707–2720

Bäumlein H, Müller AJ, Schiemann J, Helbing D, Manteuffel R, Wobus U (1987) A legumin B gene of *Vicia faba* is expressed in developing seeds of transgenic tobacco. Biol Zentralbl 106:569–575

Bäumlein H, Boerjan W, Nagy I, Panitz R, Inzé D, Wobus U (1990) Upstream sequences regulating legumin gene expression in heterologous transgenic plants. Mol Gen Genet 225:459–467

Benoist C, Chambon P (1981) In vivo sequence requirements of the SV40 early promoter regions. Nature 290:304–310

Bouman F (1975) Integument initiation and testa development in some *Cruciferae*. Bot J Linn Soc 70:213–229

Bardford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Brown JWS (1986) A catalogue of splice junction and putative branch point sequences from plant introns. Nucleic Acids Res 14:9549–9559

Bustos MM, Guiltinan MJ, Jordano J, Begum D, Kalkan FA, Hall TC (1989) Regulation of β-glucuronidase expression in transgenic tobacco plants by an A/T-rich, cis-acting sequence found upstream of a french bean β-phaseolin gene. Plant Cell 1:839–853

Colot V, Robert LS, Kavanagh TA, Goldsbrough AP, Bevan MW, Thompson RD (1987) Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. EMBO J 6:3559–3564

Croy RRD, Evans M, Narwood JN, Harris N, Gatehouse JA, Shirsat AH, Kang A, Ellis JR, Thompson A, Boulter D (1988) Expression of pea legumin sequences in pea, *Nicotiana* and yeast. Biochem Physiol Pflanzen 183:183–197

Deblaere R, Bytebier B, De Greve H, Deboeck F, Schell J, Van Montagu M, Leemans J (1985) Efficient octopine Ti plasmid-derived vectors for *Agrobacterium*-mediated gene transfer to plants. Nucleic Acids Res 13:4777–4788

Dickinson CD, Evans PR, Nielsen NC (1988) RY repeats are conserved in the 5'-flanking regions of legume seed-protein genes. Nucleic Acids Res 16:371

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to higher specific activity. Anal Biochem 132:6–13

Goldberg RB (1986) Regulation of plant gene expression. Philos Trans R Soc Lond Series B 314:343–353

Goldberg RB (1988) Plants: Novel developmental processes. Science 240:1460–1467

Goldberg RB, Barker SJ, Perez-Grau L (1989) Regulation of gene expression during plant embryogenesis. Cell 56:149–160

Heim U, Schubert R, Bäumlein H, Wobus U (1989) The legumin gene family: structure and evolutionary implications of *Vicia faba* B type genes and pseudogenes. Plant Mol Biol 13:653–663

Herman LMF, Van Montagu MC, Depicker AG (1986) Isolation of tobacco DNA segments with plant promoter activity. Mol Cell Biol 6:4486–4492

Hong GF (1982) A systematic DNA sequencing strategy. J Mol Biol 158:539–549

Horsch RR, Fry JF, Hoffmann NL, Wallroth M, Eichholtz DA, Rogers SG, Fraley RT (1985) A simple and general method for transferring genes into plants. Science 277:1229–1231

Hunger HD, Coutelle C, Behrendt G, Flackmeister C, Rosenthal A, Speer A, Breter H, Szargan R, Franke P, Stahl J, Cuong NV, Barchend G (1986) CCA paper: a new two dimensional cyanuric chloride-activated matrix for universal application in molecular biology. Anal Biochem 156:286–299

Jefferson RA (1987) Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol Biol Rep 5:387–405

Joshi CP (1987a) Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. Nucleic Acids Res 15:9627–9640

Joshi CP (1987b) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res 15:6643–6653

Krebbers E, Herdies L, De Clercq A, Seurinck J, Leemans J, Van Damme J, Segura M, Gheysen G, Van Montagu M, Vandekerckhove J (1988a) Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family. Plant Physiol 87:859–866

Krebbers E, Seurinck J, Herdies L, Cashmore AR, Timko MP (1988b) Four genes in two diverged subfamilies encode the ribulose-1,5-bisphosphate carboxylase small subunit polypeptides of *Arabidopsis thaliana*. Plant Mol Biol 11:745–759

Lee JS, Brown WE, Graham JS, Pearce G, Fox EA, Dreher TW, Ahern KG, Pearson GD, Ryan CA (1986) Molecular characterization and phylogenetic studies of a wound-inducible proteinase inhibitor I gene in *Lycopersicon* species. Proc Natl Acad Sci USA 83:7277–7281

Matzke AJM, Stöger EM, Schernthaner JP, Matzke MA (1990) Deletion analysis of a zein gene promoter in transgenic tobacco plants. Plant Mol Biol 14:323–332

Müntz K, Horstmann C, Schlesier B (1986) Seed proteins and their genetics in *Vicia faba* L. Biol Zentralbl 105:107–120

Peleman J, Boerjan W, Engler G, Seurinck J, Botterman J, Alliotte T, Van Montagu M, Inzé D (1989) Strong cellular preference in the expression of a housekeeping gene of *Arabidopsis thaliana* encoding *S*-adenosylmethionine synthetase. Plant Cell 1:81–93

Perez-Grau L, Goldberg RB (1989) Soybean seed protein genes are regulated spatially during embryogenesis. Plant Cell 1:1095–1109

Pustell J, Kafatos FC (1986) A convenient and adaptable micro-

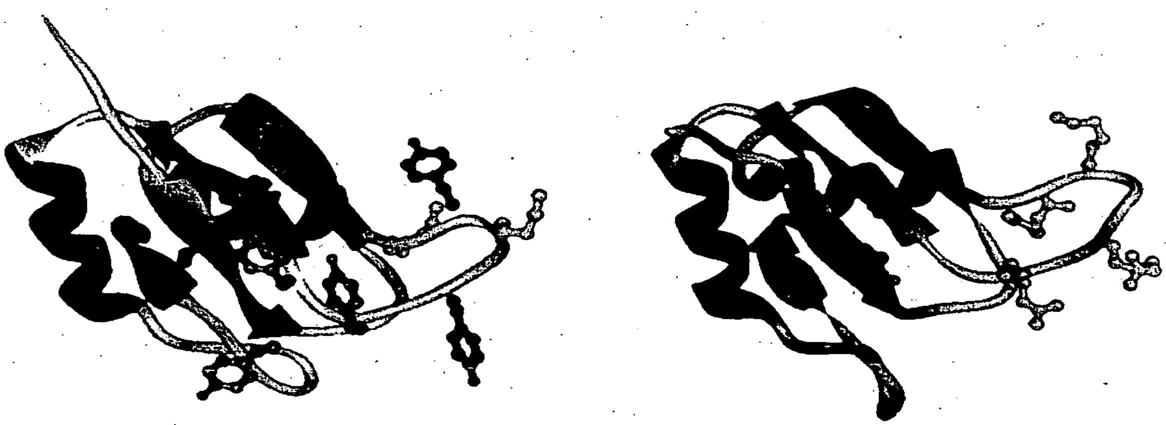
computer environment for DNA and protein manipulation and analysis. Nucleic Acids Res 14:479–488
Reiss B, Sprengel R, Will H, Schaller H (1984) A new sensitive method for quantitative and qualitative assay of neomycin phosphotransferase in crude cell extracts. Gene 30:211–218
Sanger F, Nicklen S, Coulson AR (1977) DNA-sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467
Valvekens D, Van Montagu M, Van Lijsebettens M (1988) *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis* root explants using kanamycin selection. Proc Natl Acad Sci USA 85:5536–5540
Vaughan JG, Whitehouse FLS, Whitehouse JM (1971) Seed structure and the taxonomy of the *Cruciferae*. Bot J Linn Soc 64:383–409
Weschke W, Bäumlein H, Wobus H (1987) Nucleotide sequence of a field bean (*Vicia faba* L. var minor) vicilin gene. Nucleic Acids Res 15:10065
Wobus U, Bäumlein H, Bassüner R, Heim U, Jung R, Müntz K, Saalbach G, Weschke W (1986) Characteristics of two types of legumin genes in the field bean (*Vicia faba* L. var. minor) genome as revealed by cDNA analysis. FEBS Lett 201:74–80
Wobus U, Bäumlein H, Inzé D, Nagy I (1989) *Vicia faba* storage protein genes and their promoter activity in transgenic tobacco plants. In: Galling G (ed) Applied plant molecular biology. Technical University, Braunschweig, pp 98–103

Communicated by J. Schell

mber 1997

JMB

Journal of Molecular Biology



Academic Press

272 (1) 1–165 ISSN 0022–2836

0022-2836(199709)272:1;1-R

JMB

Strategies for the *in vitro* Evolution of Protein Function: Enzyme Evolution by Random Recombination of Improved Sequences

Jeffrey C. Moore, Hua-Ming Jin, Olga Kuchner and Frances H. Arnold*

Division of Chemistry and Chemical Engineering, Mail Code 210-41, California Institute of Technology Pasadena, CA 91125, USA

**Corresponding author*

Sets of genes improved by directed evolution can be recombined *in vitro* to produce further improvements in protein function. Recombination is particularly useful when improved sequences are available; costs of generating such sequences, however, must be weighed against the costs of further evolution by sequential random mutagenesis. Four genes encoding para-nitrobenzyl (pNB) esterase variants exhibiting enhanced activity were recombined in two cycles of high-fidelity DNA shuffling and screening. Genes encoding enzymes exhibiting further improvements in activity were analyzed in order to elucidate evolutionary processes at the DNA level and begin to provide an experimental basis for choosing *in vitro* evolution strategies and setting key parameters for recombination. DNA sequencing of improved variants from the two rounds of DNA shuffling confirmed important features of the recombination process: rapid fixation and accumulation of beneficial mutations from multiple parent sequences as well as removal of silent and deleterious mutations. The five to sixfold further enhancement of total activity towards the para-nitrophenyl (pNP) ester of loracarbef was obtained through recombination of mutations from several parent sequences as well as new point mutations. Computer simulations of recombination and screening illustrate the trade-offs between recombining fewer parent sequences (in order to reduce screening requirements) and lowering the potential for further evolution. Search strategies which may substantially reduce screening requirements in certain situations are described.

Keywords: directed evolution; DNA shuffling; random mutagenesis; para-nitrobenzyl esterase

Introduction

Enzymes can be evolved *in vitro* to exhibit new and useful functions. A sampling of the local sequence space of the enzyme is created by mutagenesis; screening or selection directs the evolution towards the desired features. A successful strategy for improving enzyme activity in non-natural environments (Chen & Arnold, 1993) and on non-natural substrates (Moore & Arnold, 1996) has been to accumulate amino acid substitutions over multiple generations of random mutagenesis and screening. In practice, the best variant identified in each generation is chosen to parent the subsequent generation. Other potentially useful variants are set aside, and their mutations must be rediscovered in the evolved protein background in order to become incorporated. Because there is no mechanism other than back mutation for deleting mutations, this approach can also accumulate deleterious mutations, leading to premature termination of an evolving lineage. These are the classical arguments for the benefits of recombination (sex) in evolution (Maynard Smith, 1988). Recombination allows more rapid accumulation of beneficial mutations present in a population. It also makes possible the removal of deleterious mutations which would otherwise accumulate in an asexual population, a phenomenon known to geneticists as Müller's ratchet (Müller, 1932). Recombination can

Present addresses: J. C. Moore, Merck & Co., Inc., P.O. Box 2000, RY80Y-110, Rahway, NJ 07065, USA; H.-M. Jin, SmithKline Beecham, Mail Code UE0548, 709 Swedeland Rd., King of Prussia, PA 19406, USA.

Abbreviations used: pNB, para-nitrobenzyl; pNP, para-nitrophenyl.

0022-2836/97/380336-12 $25.00/0/mb971252

provide similar benefits for *in vitro* molecular evolution (Stemmer, 1994a,b).

Bacillus subtilis p-nitrobenzyl (pNB) esterase catalyzes the hydrolysis of the para-nitrobenzyl esters of various cephalosporin-type antibiotics, a necessary step in their large-scale synthesis (Zock *et al.*, 1994). Using four generations of sequential random mutagenesis and screening, we evolved a series of pNB esterases up to 30 times more active towards hydrolysis of the pNB ester of loracarbef (LCN-pNB) in aqueous dimethylformamide (Moore & Arnold, 1996). During the fourth generation, a large number (~7500) of pNB esterase clones were screened and partially characterized in order to validate the rapid screening assay. Sixteen improved pNB esterase clones were identified, from which the five most active enzymes (>50% enhancements in activity over the parent enzyme) were characterized. DNA sequencing revealed four unique pNB esterases (Table 1). Due to the limitations of screening, evolved sequences are generated using a low rate of point mutagenesis and typically accumulate a single beneficial mutation per generation. A simple restriction/ligation experiment demonstrated that recombination of mutations present in at least two of those sequences could further improve pNB esterase activity. Recombining gene segments from two improved pNB esterase variants yielded an enzyme twice as active as the best parent. DNA sequencing demonstrated that mutations from each of the two parents were combined in the new sequence (I60V and L334S), while one neutral or slightly deleterious mutation was deleted (K267R; Moore & Arnold, 1996).

Stemmer recently introduced the technique of "DNA shuffling" to create novel genes by recombination of closely-related DNA sequences (Stemmer, 1994b). Because it also introduces new point mutations during reassembly of the DNA fragments, DNA shuffling alone has been effective for directed protein evolution starting from a single sequence (Stemmer, 1994a; Crameri *et al.*, 1996). Questions arise as to how this approach is best implemented and integrated with other *in vitro* evolution approaches such as sequential random mutagenesis. Issues include optimizing the point mutagenesis rate associated with DNA shuffling, determining appropriate screening sample sizes and how many parental genes to recombine, and deciding when to use recombination. Here we investigate the further evolution of pNB esterase by DNA shuffling of the improved sequences generated by random mutagenesis and screening. By following how the genes evolve during cycles of DNA shuffling and screening, we can elucidate the mechanisms contributing to the evolution of function and begin to optimize strategies for *in vitro* evolution. An analysis of the recombination process identifies some of its benefits and limitations for directed evolution and allows a rational choice of mutagenesis and screening strategies.

Results and Discussion

Recombination statistics and screening requirements

To comment on the utility of DNA shuffling in directed evolution, a review of the statistics of recombination of multiple parent sequences is useful. For this discussion, we will assume that the mutations are unique and distributed far enough from one another on the genes that recombination occurs freely between any two. Furthermore, equal amounts of the initial DNA sequences are recombined. Consider the random recombination of three parent sequences, each of which contains a single mutation. Any given mutation will be incorporated into a progeny sequence with a probability of 1/3; the probability of generating the wild-type sequence is 2/3 at each mutation site. This highlights an important consequence of shuffling multiple sequences: there is a statistical preference for the absence of mutation in the progeny. The overall probability of picking a completely wild-type sequence from the recombined library is $(2/3)^3 = 0.30$. The probability of generating a sequence containing a single mutation (a parent sequence) is $1/3 \times (2/3)^2 = 0.15$. Because there are $C_1^3 = 3!/1!2!$, or three such sequences, the overall fraction of parent sequences in the library is 0.45. Thus fully 75% of the sequences in the recombined library are variants already in the evolutionist's possession.

In general, for a recombination system consisting of N sequences and M total mutations, the probability of generating progeny sequences containing μ mutations equals the number of ways a μ-mutation sequence can be generated (C_μ^M) multiplied by the probability of generating any single μ-mutation sequence:

$$P_\mu = C_\mu^M \left(\frac{1}{N}\right)^\mu \left(\frac{N-1}{N}\right)^{M-\mu}$$
$$= \frac{M!}{(M-\mu)!\mu!} \left(\frac{1}{N}\right)^\mu \left(\frac{N-1}{N}\right)^{M-\mu}.$$

Figure 1 summarizes the analysis for recombination of single-mutation parent sequences ($N = M$). The probability that recombination will return the zero-mutation "grandparent" or single-mutation parent sequences remains constant between 73 and 75%; only ~25% of the clones screened have sequences that have not already been examined. The probability of creating individual sequences declines dramatically with increasing numbers of parents. The least frequent sequences are those containing the majority of mutations from the parent population, and the sequence containing all the mutations ($\mu = M$) is of course the rarest. The probability P_M of generating the rarest sequence is $1/N^M$.

Because we are interested in the evolution of function, we need consider only those mutations responsible for functional differences among pro-

Table 1. DNA and amino acid substitutions in fourth, fifth and sixth generation evolved pNB esterases

Mutation	Amino acid substitution	4-54B9	4-38B9	4-53D5	4-43E7	5-6C8	5-5E4	5-4H4	5-4G2	5-4D12	5-2D3	6-10F1	6-1D12	6-1C7	6-1A6
ATC 27 → ATA									YES	yes					
CCT 33 → CCC		yes													
ATT 60 → GTT	I60V		yes			yes	yes		yes	yes	yes	yes	yes	yes	yes
GAT 81 → GAC															yes
TAT 84 → TAC				yes		yes									
AGT 94 → GGT	S94G				yes										
GGA 127 → GGG															yes
TCG 148 → TCC						yes									
TTT 149 → TTC									yes	yes					
GGC 227 → GGG	A227G										yes				
ATT 239 → ATC				yes			yes								
AGA 246 → AGG															yes
GCG 252 → GCT							yes								
AAA 267 → AGA	K267R	yes													
CCG 317 → TCG	P317S					yes						yes	yes	yes	yes
TTA 334 → GTA	L334V				yes										
TTA 334 → TCA	L334S	yes				yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
GCT 343 → GTT	A343V			yes				yes							yes
CAT 356 → CGT	H356R								yes	yes		yes		yes	
ACT 359 → GCT	T359A												yes		
ATT 464 → GTT	I464V								yes	yes		yes		yes	

DNA substitutions are identified in the context of the three-base codon of the encoded enzyme sequence. Grey background indicates mutations from fourth generation parent sequences. White background indicates new mutations which arose during DNA shuffling. Bold face type indicates translated DNA mutations.

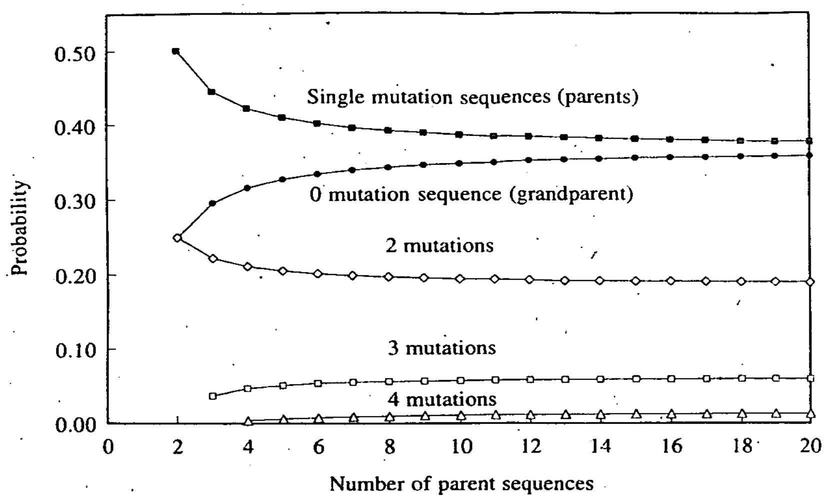


Figure 1. Probabilities of generating sequences containing different numbers of mutations by random recombination, based on recombining single-mutation parent sequences. Novel variants (not grandparent or parent sequences) are shown with unfilled symbols.

tein variants. Neutral mutations by definition do not affect function; their distribution among progeny sequences is determined statistically, even in the screened population (Zhao & Arnold, 1997b). Thus for the purposes of this discussion of recombination libraries and screening requirements, M is the number of mutations that affect the targeted function (either beneficial or deleterious).† By screening enough clones to ensure that the rarest sequence, that is, containing all M mutations, has been examined, one can be sure that the best variant will be discovered. This is true even if the best variant does not contain all the functional mutations (as would be expected if some mutations were deleterious or if the effects of mutations are not cumulative).

In practice, of course, oversampling is required to ensure that a particular variant has been examined during the course of screening. To be 95% confident that the most active combination variant has been examined, we must be 95% confident the rarest variant has been examined. If S is the number of clones sampled, then

$$(1 - P_M)^S < 1 - \text{confidence limit}$$

describes how the probability of not sampling the rarest variant changes with increasing S. This allows calculation of the number of samples required for a given confidence limit. The oversampling is then how many more samples must be screened over the theoretical minimum. When one clone is required with 95% confidence, the oversampling will be between 2.6 and 3.0 (for larger numbers of parents). Even a relatively low rate of background point mutagenesis, however, can introduce significant confounding effects. Non-neutral point mutations obscure recombination events and increase the amount of screening required to find the best sequences (*vide infra*). Thus, in practice, it may be impossible to screen sufficient numbers of clones to be sure of finding the best recombinant, particularly when the point mutation rate is high and a large number of functional mutations are being recombined. Alternative strategies which can reduce screening requirements under special conditions will be discussed further on.

† A mutation that is neutral in one context (i.e. in the wild-type background), but becomes functional in a different context, would be considered a functional mutation.

DNA shuffling of evolved pNB esterases

An effect of forcing DNA polymerase to synthesize full length genes from the pool of small DNA fragments generated during DNA shuffling is additional background point mutagenesis. A high rate of point mutagenesis can severely inhibit the discovery of novel combinations of existing mutations within a population. Because most mutations are deleterious (in a screening assay sensitive to small changes in the screening variable), beneficial recombinations and rare beneficial point mutations are masked by the negative background. DNA shuffling with a 0.7% mutagenesis rate, for example, would yield an average of 10-11 point mutations in the 1470 bp pNB esterase gene. This is substantially more than the optimal mutation frequency (~three mutations per gene) for directed evolution of pNB esterase (Moore & Arnold, 1996). In fact, when the four evolved pNB esterase gene sequences were shuffled using *Taq* polymerase, fully 90% of the clones in the resulting library exhibited essentially no esterase activity during screening (data not shown). In a parallel study, we observed that 80% of the clones generated by DNA shuffling of subtilisin E exhibited no activity (Zhao & Arnold, 1997a).

In an effort to reduce the background mutagenesis rate, a proofreading polymerase (Pwo) was used during fragment reassembly. With Pwo, 50 to 100 base-pair fragments could be reassembled to create a library in which fully 80% of the clones

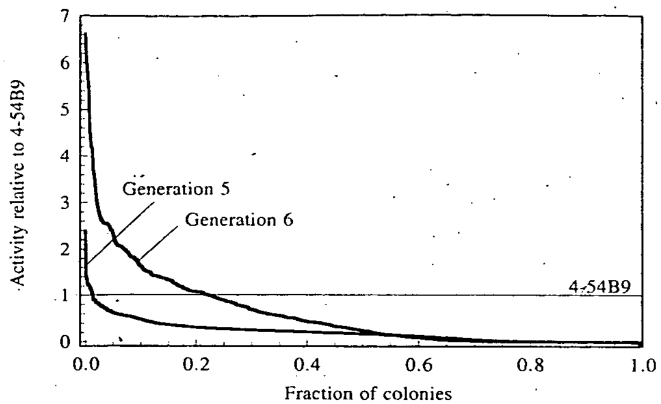


Figure 2. Activity profiles of generations 5 and 6 determined by screening libraries created by DNA shuffling of unique fourth and fifth generation variants. Activities were sorted from best to worst. Profiles are normalized by the number of clones screened.

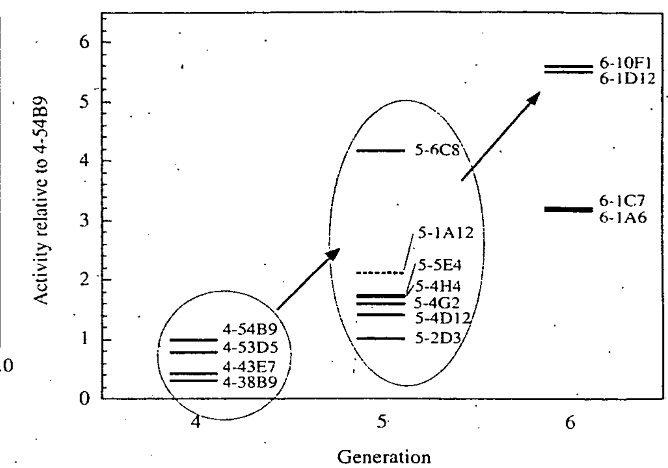


Figure 3. Activities of fourth, fifth and sixth generation pNB esterase variants (Table 1) in screening assay. Fourth generation variants were recombined and screened to identify improved enzymes in generations 5 and 6.

retained activity. Inserts from 13 randomly picked colonies were partially sequenced in order to determine the point mutation rate. Five mutations not present in any of the parent sequences were found in 12,000 nucleotides sequenced, for an overall mutagenic rate of ~0.04%. These minimally mutagenic conditions were used for DNA shuffling. A subsequent, in-depth investigation of the various steps involved in DNA shuffling has allowed us to identify a set of recombination protocols with a wide range of point mutagenesis rates (Zhao & Arnold, 1997a).

Four unique fourth generation improved pNB esterase variants were chosen as the starting point for further directed evolution by DNA shuffling. Two cycles of DNA shuffling and screening for activity towards the *p*-nitrophenyl ester of loracarbef (pNP-LCN) in 25% dimethylformamide (DMF) were performed. The activity profiles of the resulting populations (generations 5 and 6) are shown in Figure 2. To generate these profiles, activities of the individual clones measured in the 96-well plate screening assay were normalized by cell density (A_{600}) and plotted in descending order. Approximately 2% of the 948 generation 5 clones screened exhibit more total activity than the most active parent (4-54B9). The screened population was sufficiently large to give a high level of confidence that the most active variant that can be generated by simple recombination of the fourth generation sequences has been found.† The six most active variants from generation 5 were collected and shuffled again to create generation 6. Fully 20% of the 474 clones screened were more active than 4-54B9. Only 20 to 25% of the clones were inactive, as expected using the high fidelity Pwo-only shuffling conditions.

Figure 3 summarizes the activities of the four fourth generation parents and the best variants identified in generations 5 and 6. The improvement in enzyme activity as a result of shuffling is already apparent in the fifth generation, which includes one variant (5-6C8) fourfold more active than 4-54B9 and twice as active as variant 5-1A12 previously generated by ligation recombination (Moore & Arnold, 1996). The sixth generation contains two clones with yet higher activities than 5-6C8. The best one, 6-10F1, represents a five to sixfold improvement over 4-54B9 and is ~150 times more active than the wild-type.

Activities of the fifth and sixth generation variants towards the *p*-nitrobenzyl ester of loracarbef (LCN-pNB) were also determined, using a modified HPLC assay as described in Materials and Methods. The best pNB esterase is 5-6C8, which exhibits a threefold increase in total activity over 4-54B9. This clone is now ~100 times more active than wild-type pNB esterase towards LCN-pNB in 25% DMF. The sixth generation variants exhibited no further improvement in activity towards this substrate, a clear reflection of the use of the pNP ester during screening and the first law of random mutagenesis: "You get what you screen for" (You & Arnold, 1996).

† When shuffling four parent sequences each of which contains one beneficial mutation, 765 clones must be screened to be 95% confident that all combinations have been examined (assuming recombination occurs freely between mutations and no point mutagenesis). A 0.04% rate of point mutagenesis translates to less than 0.6 new mutations per sequence, of which only a fraction will affect function (estimated from the activity profile of a library created by error-prone PCR to be ~0.5, data not shown).

Analysis of evolved pNB esterase genes

DNA mutations present in the four parent fourth generation sequences and mutations identified by sequencing the genes encoding the selected fifth and sixth generation variants are summarized in Table 1. By comparing the activities and sequences of these variants with the third-generation parent, four beneficial mutations were identified (leading to amino acid substitutions I60V, L334V, L334S and A343V). The remaining mutations present in the fourth generation sequences are neutral or mildly deleterious (Moore & Arnold, 1996).

Several interesting observations can be made from this Table. It can be seen that a number of mutations increase their frequencies in the subsequent generations. Substitutions I60V in 4-38B9 and L334S in 4-54B9 are each present in a single fourth generation parent. In contrast, I60V is present in five of the six fifth-generation variants, and L334S is present in all six. By the sixth generation both substitutions are fixed in the population. A new substitution at position 317, first found during the fifth generation (5-6C8), also becomes fixed by the sixth. This new mutation probably accounts for the significant increase in activity of variant 5-6C8. The P317S substitution is positioned near the enzyme surface in a loop located on the same side of the entrance to the substrate binding pocket as amino acid substitutions L334S, M358V and A343V (Moore & Arnold, 1996). Removal of a proline at this position may relax conformational constraints on the loop, allowing the substrate freer access to the active site.

The two separate beneficial mutations at position 334 in 4-43E7 and 4-54B9 are mutually exclusive, and a competition exists as to which one will be propagated to successive generations. Variant 4-54B9 has more than twice the activity of 4-43E7 as a result of the mutation at position 334, and the fifth generation recombination progeny in fact show the L334S substitution from 4-54B9 exclusively. Recombination provides a rapid means to identify the most effective mutation among multiple possibilities at any given site.

Related to the observation that beneficial mutation combinations are fixed is the fact that recombination and screening also effectively remove neutral and deleterious mutations. Three of the five mutations present in the fourth generation parents that are synonymous (DNA mutations in codons 33, 84, and 239 that do not lead to amino acid substitutions) or non-synonymous, but believed neutral or mildly deleterious in their effects on total activity (mutations leading to amino acid substitutions S94G and K267R (Moore & Arnold, 1996)), have been removed from the improved pNB esterase population in a single round of shuffling; all five are removed by the sixth generation. The two most active sixth generation enzyme variants, 6-10F1 and 6-1D12, have no synonymous mutations at all and only one mutation (at position 359) not seen in any previous clone. Due to the statistical preference for the absence of mutations the recombination process is highly effective in filtering out neutral (and deleterious) mutations starting from multiple parent sequences.

Table 1 also shows that the DNA shuffling technique can recombine multiple parent sequences to create novel progeny. Recombination between at least three fourth-generation parents is required to create 5-5E4, and at least three fifth-generation parents were recombined to generate clones 6-10F1 and 6-1A6 (based on the presence and absence of the DNA mutations in the sequences compared to the parent sequences).

Finally, it is useful to note that DNA shuffling generates point mutations that are rarely observed during PCR (at least for the low-mutagenesis rate PCR conditions used for directed evolution of longer DNA sequences). Four of the 12 new point mutations identified in the fifth and sixth generation variants, for example, are $G \rightarrow C$ (and $C \rightarrow G$) and $G \rightarrow T$ (and $C \rightarrow A$) transversions, which were not found at all during the first four generations of pNB esterase evolution involving PCR mutagenesis (Moore & Arnold, 1996). These mutations were also generated very rarely during the error-prone PCR mutagenesis of subtilisin (Shafikhani *et al.*, 1997). DNA shuffling and error-prone PCR together may provide access to a wider range of amino acid substitutions.

Evolved pNB esterase amino acid sequences

Amino acid substitutions in the evolved pNB esterases are indicated in Table 1; changes in amino acid sequence along the lineage are summarized in Figure 4. The accumulation and fixation of two beneficial amino acid substitutions from the fourth generation, I60V and L334S, is essentially complete in a single generation of DNA shuffling and screening 948 clones. In contrast, A343V, a beneficial mutation found in the fourth generation, no longer appears in the majority of fifth or sixth generation variants. The (5-4H4) recombinant of the parent containing this mutation (4-53D5) with 4-54B9 shows no improvement in activity over 4-54B9 (Figure 3). Substitutions A343V and L334S therefore do not work in concert to improve enzyme activity, and consequently there is little or no driving force to retain A343V in the population. The remaining fifth generation variants, with the exception of 5-6C8, are less active than 5-1A12 (Figure 3), yet they contain the I60V and L334S substitutions while omitting K267R, as does 5-1A12. This suggests that the additional mutations found in those sequences are neutral, or possibly, deleterious. For instance, the amino acid sequences of 5-5E4 and 5-1A12 are identical, and the decreased activity of the former is likely due to the two synonymous mutations in 5-5E4 not present in 5-1A12. Because the screen evaluates the total activity of a clone (normalized by cell density), synonymous mutations can influence the result, for

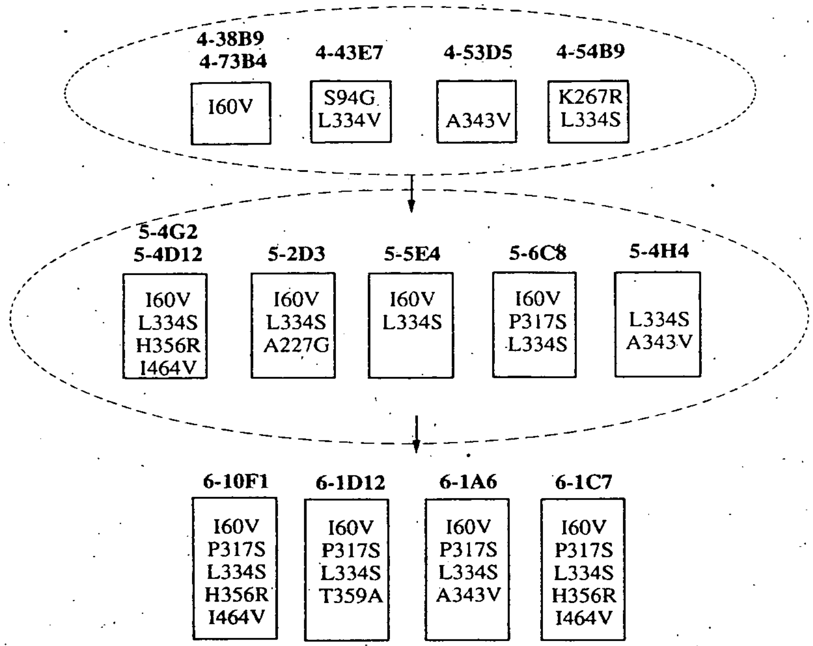


Figure 4. Lineage of pNB esterase variants showing amino acid substitutions accumulated by four generations of sequential random mutagenesis (fourth generation) and by DNA shuffling (fifth and sixth generations) and screening. All variants contain amino acid substitutions H322R, Y370F, M358V and L144M from the third generation parent (Moore & Arnold, 1996).

example, by affecting the amount of active enzyme expressed. The new beneficial mutation that gives rise to the P317S substitution becomes fixed in the sixth generation, and further evolution during that generation primarily arises from point mutation rather than recombination.

Clones 5-4G2 and 5-4D12, whose DNA sequences are identical, both contain amino acid substitutions H356R and I464V. These two substitutions are seen together again in 6-10F1 and 6-1C7. Because 6-10F1 and 6-1D12 have almost identical activity, we can reasonably infer that the I60V, P317S, and L334S substitutions are responsible for that activity, while the mutations leading to H356R and I464V from the fifth generation as well as a new mutation, T359A, in 6-1D12 are neutral. The three mutations believed responsible for enhanced activity are also present in 6-1A6, along with the last mutation in this system known to enhance activity, A343V. That 6-1A6 has lower activity than 6-10F1 and 6-1D12 is therefore attributable to either the three synonymous mutations in 6-1A6 (Table 1) or antagonism between amino acid substitutions A343V and P317S or I60V.

The new point mutations that arose during the minimally mutagenic DNA shuffling increased (P317S) and decreased enzyme activity. The effects of individual mutations can be ascertained with confidence because the sequences differ from one another at very few positions. We have recently demonstrated a method that allows one to distinguish clearly beneficial, neutral and deleterious mutations in evolved sequences by random recombination with ancestor sequences (Zhao & Arnold, 1997b). This method will be particularly useful for identifying mutations responsible for functional changes in proteins in a background of neutral mutations (as happens when multiple new mutations are present).

Only 2% of the fifth generation clones are more active than the most active parent, 4-54B9 (Figure 2). Although 25% of the progeny should be novel, the combination I60V + L334S predominates in the most active variants (Figure 4), suggesting that many of the remaining combinations lead to lower activity than in 4-54B9. Additionally, while there is no mechanism for recombination alone to generate inactive clones, ~25% of the variants in Figure 2 are inactive, presumably as a result of background point mutation. This implies that the frequency of enhanced-activity recombinants is reduced by point mutation and emphasizes the importance of minimizing the mutagenesis rate when recombining positive mutations.

Developing strategies for directed evolution

Recombination versus *random mutagenesis*

Recombination is only useful if a population of sequences is available from which new combinations of mutations can be generated. Homologous proteins with similar sequences could provide such a starting population (Stemmer, 1994b). (Note, however, that a high level of sequence identity may be required for DNA shuffling.) Populations of sequences can also be created by the background point mutagenesis feature of DNA shuffling (Crameri *et al.*, 1996). Alternatively, they can be generated by random mutagenesis and screening experiments, as they have been for the current study. When interesting sequences already exist, recombination offers an efficient means to use that information. If the sequences must be generated, however, then one should consider that

cost in the overall cost of evolution by recombination as compared to, for example, evolution by sequential generations of random mutagenesis and screening.

In theory, the sequential (or "asexual") approach requiring the least labor in terms of screening is to screen randomly mutagenized clones until a positive is identified and then use that as the template for the next generation. The process is a random walk in which the first uphill step encountered is taken. To take a simple illustration, consider three mutations A, B and C that each contribute in a cumulative, if not additive, manner when combined. A, B and C could be collected in the ABC variant in three sequential generations of mutagenesis and screening. Alternatively, if A, B and C all contribute to the desired feature in the wild-type background (as they often do; see, for example, Chen & Arnold, 1993), they could be found separately and then recombined to make ABC. Finding the single-mutation sequences A, B, and C, however, requires screening the same number of colonies as finding ABC by sequential evolution. Recombining the A, B, and C sequences to make ABC requires additional screening. Of course, the sequential pathway requires three random mutagenesis steps, while the recombination pathway requires only one mutagenesis step and one DNA shuffling step. The advantages of one approach over the other then depend on the costs of screening relative to the DNA manipulations.

Note that the severe limitations screening places on the number of colonies that can be sampled makes it difficult to accept downhill steps in the hope that further improvements can be found further out in sequence space (Moore & Arnold, 1997). It also means that extremely rare events such as the recombination of neutral or slightly deleterious mutations to make a beneficial combination will probably not contribute in any significant fashion to the evolutionary process.

The pNB esterase evolution provides a concrete example for analysis. Approximately one in every 1500 to 2000 randomly mutagenized pNB esterase clones screened was positive (showing 50% or greater enhancement in activity over the parent; Moore & Arnold, 1996). To generate the population of four unique positives for DNA shuffling, we examined a total of 7500 clones. Finding the best combination variant required additional DNA shuffling experiments, and ~1400 additional colonies were screened. Thus a total of 9000 clones were screened in going from generations 3 to 6. There is no guarantee that the sequences chosen for recombination are unique: in fact, the original fourth generation clones contained five variants, two of which were identical (4-38B9 and 4-54B9) and two of which contained mutations in the same codon (4-43E7 and 4-54B9), precluding recombination between these variant pairs. It is very likely that variants of comparable or even greater activity could also have been created by continuing random mutagenesis and screening for three generations from the first fourth generation variant identified. The total screening requirement would be the same.

In practice, however, the uphill climb often involves identification of multiple positives during each generation. Everything but the one chosen to parent the next generation is discarded in the random uphill walk of the "asexual" evolution. During the pNB esterase evolution, we often identified four or five potential positives during the rapid screen on the LCN-pNP colorimetric substrate. Those were either verified or not during a second level screen on the *p*-nitrobenzyl (LCN-pNB) substrate, and it was often the case that more than one sequence was a true positive (Moore & Arnold, 1996). The other improved sequences could of course be collected and recombined at any time and at relatively little screening cost. A significant advantage of the DNA shuffling method is its ability to utilize these available positive sequences.

Computer simulations of random recombination and screening

The statistical model can be used to optimize the number of parent sequences chosen for DNA shuffling. Screening during the fourth generation actually resulted in the identification of 16 clones measurably more active than the parent, of which five were at least 50% more active (Moore & Arnold, 1996). An attempt to recombine all 16 sequences yielded no clones more active than 4-54B9 (~1000 clones screened). This result can be understood when we consider the dramatically lower probability of finding the best combination(s) as the number of sequences increases. If the screening sample size is limited to a few thousand clones, there is little chance that the best sequences, or even sequences better than the best parent, will be found by screening a library created from 16 parents.

We have used a computer simulation of the random sampling of the two recombined libraries obtained by shuffling five and ten sequences to illustrate the advantage of choosing fewer parents when screening is limited. Recombining all ten parents becomes advantageous, however, when large numbers of clones can be examined. (Of course, the larger sampling requirement should then be compared to the potential for continued evolution by random mutagenesis.) Assuming that the ten parent sequences each contain a unique, single beneficial mutation ($N = M$) and that they can be recombined to give all possible combinations, we calculated P_μ for $\mu = 0$ through 10. Since $\Sigma P_\mu = 1$, these were organized into a cumulative distribution from 0 to 1, and a random number generator was used to pick a point on the cumulative distribution, thereby identifying μ (number of mutations per sequence). A second random number generator was used to pick one of the C_μ^M possible sequences containing μ substitutions using an evenly spaced distribution of possible

combinations. The activity of the sequence chosen was then calculated by assuming that the free energies of activation of the variants (proportional to the natural logarithms of their activities) are additive.

The results of this simulation are shown in Figure 5, using the activity data from the fourth generation pNB esterase variants. Figure 5(a) shows the averages of the highest values of mutant activities obtained over 15 separate trials for each (screening) sample size. The results obtained by shuffling the ten best mutants (black diamonds) can be seen to be slightly worse than those obtained by shuffling the five best mutants (white squares), for sample sizes up to about 10,000 to 15,000. That is, the average expected best mutant is higher for shuffling five parents at a time for small sample sizes. Figure 5(b) and (c) show the range of values of the highest mutant activity obtained on each of 15 separate trials for each sample size. Here, the highest values obtained from recombining the best ten variants (black diamonds) become better than the values obtained from shuffling the best five (white squares) at sample sizes greater than about 1000. Although shuffling the top ten mutants for this set of data can yield higher final activities, the simulation shows that the outcome is much more risky when screening capabilities are limited to a few thousand clones.

Simulations also show that the results of the comparison of shuffling five *versus* ten parents is highly sensitive to the values of the activities. For instance, if the activities of mutants 6 through 10 are decreased, then the sample size at which recombining all ten mutants becomes preferable becomes much higher. Moreover, the simulation can be adapted for cases in which some or all of the parent sequences have two or more mutations, which may or may not be recombinable. Thus this simulation approach can be used to determine the optimal number of sequences to recombine for any given set of activity values and any given sample size.

The simple additivity assumption on which these simulations are based† is a reasonable first approximation of the behavior of combined mutations in proteins (Wells, 1990) and is useful for a first exploration of strategic issues in *in vitro* protein evolution. The real behavior is often more complex and will depend on the property of interest as well as the particular protein. However, it is likely that deviations from simple additivity are governed by non-linear functions of the number and magnitude of changes; values will certainly depend on which subset of mutations is recombined. While it is possible to modify the simulation to take into account deviations from additivity, very little data are available on the effects of large numbers of mutations. We have therefore not

† Both beneficial and deleterious mutations can be accommodated in this framework.

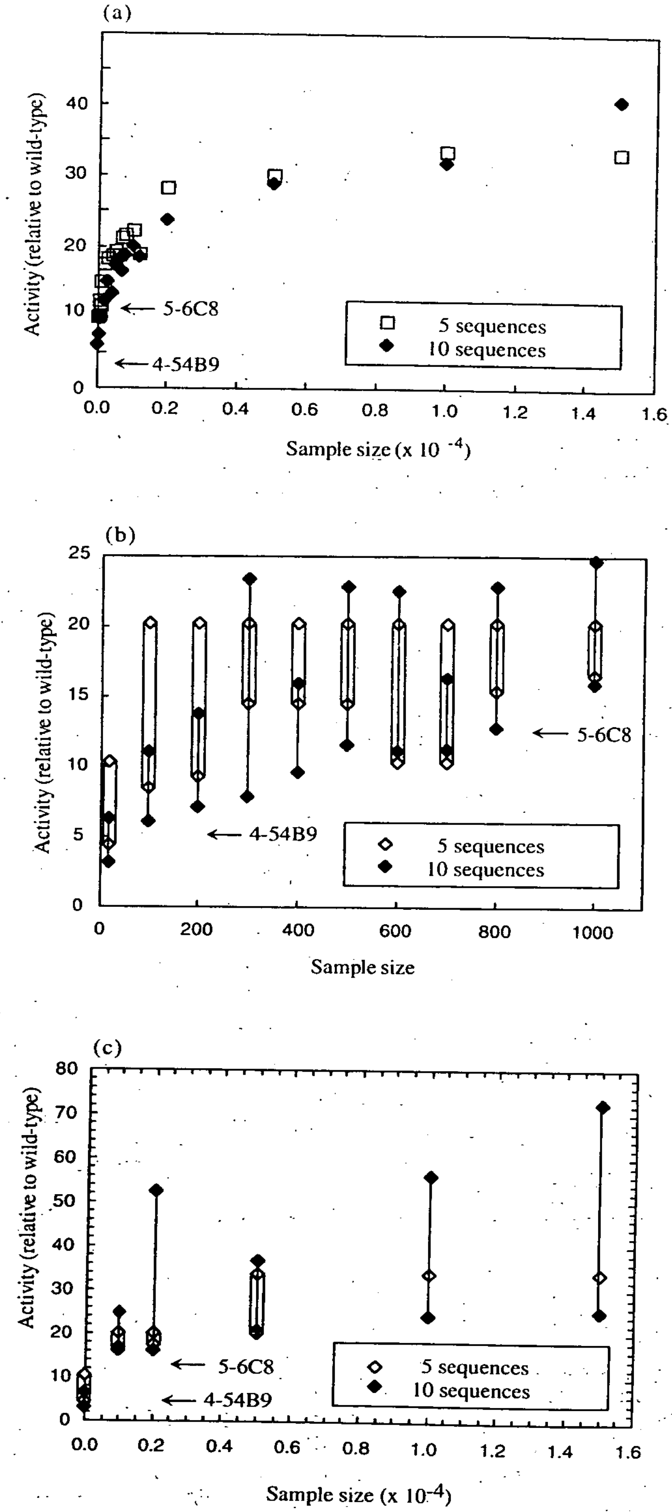


Figure 5. (a) Averages of highest values of mutant activities obtained over 15 separate trials of simulated random recombination of five and ten parent sequences. (b) and (c) Range of values of mutant activities obtained over 15 separate trials. Activities of best fourth-generation parent (4-54B9) and highest-activity fifth generation clone identified (5-6C8) are indicated for comparison.

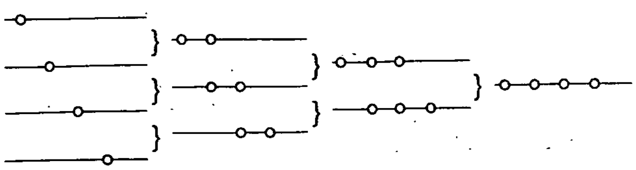


Figure 6. Pairwise recombination can reduce screening requirements, provided effects of mutations are cumulative. By shuffling two sequences at a time, sequences containing two mutations represent 25% of the recombined library. This example involves six recombination experiments.

attempted to include deviations from additivity in the current simulations. Figures 5(a), (b), and (c) show the activities of the best fourth generation parent (4-54B9) and the best fifth generation clone identified (5-6C8) by screening the shuffled library. That the activity of 5-6C8 is ~twofold less than the average expected for screening 948 clones reflects the fact that (i) only four of the original five positive clones identified during generation 4 were unique, (ii) two mutations were on the same codon and could not be recombined, and (iii) the mutations combine with significantly less than 100% additivity.

Alternative search strategies

Finally, we will briefly consider two other search strategies that might be used to minimize screening requirements. One approach to producing a multiple-mutation variant which requires the screening of far less clones is multiple-step pairwise recombination. This strategy is illustrated in Figure 6 for the simple case of recombining four (beneficial) mutations from four separate parents. Pairs of parents are mated. As each progeny is a double mutant 25% of the time, only 12 clones are required to find all the double mutants, assuming the effects of the mutations are cumulative. The double mutants are then similarly mated, and screening only eight clones will identify the triple mutants. Mating and screening four clones will generate the quadruple mutant. Thus a total of only 62 clones (24 × 2.6 times oversampling to be 95% confident at each step) must be screened, as compared to the 765 required to generate the quadruple mutant in a single recombination step. Such an approach requires considerable DNA manipulation and would be most useful when screening is extremely difficult. (An attractive alternative at this point may be sequencing the parents and recombination by site-directed mutagenesis.) A further cost of this approach is that the search space is very limited. The assumption is that each activity-enhancing mutation will contribute to the overall activity, so that the quadruple mutant is the best performer of this population. If a particular double or triple mutant is the best performer, it may or may not be found, since not all of these intermediate mutants will have been examined.

A compromise method that works well, at least in theory, can be described as "population recombination." The idea is to shuffle all four parent sequences at once and screen enough clones to see all the double mutants. Because each double mutant occurs 3.5% of the time, 28 clones must be screened. This examines all of the pair-wise interactions between mutations and eliminates those which are not cumulative. The double mutant population is recombined to produce all of the triple mutants and the quadruple mutant (requires screening 16 clones). If the mutations were at least cumulative in their effects, screening 132 (44 × 3.0 times oversampling) clones would search the space completely for the best (quadruple) mutant. This approach most closely describes how recombination/selection experiments operate (Stemmer, 1994a) where all of the clones that survive a particular selection criterion are recombined (often 100 clones or more serving as the parent population for the next generation).

Conclusions

Recombination is an important tool for directing the evolution of proteins. Beneficial mutations can be recombined, while neutral and deleterious mutations are eliminated. The need to screen rather than select for many important enzyme functions, however, severely limits the ability to search for useful combinations. It is therefore imperative to analyze various recombination strategies. Mutagenic rates associated with the recombination process must be low so that beneficial mutations are not lost in a background of deleterious ones. Although a new beneficial amino acid substitution was found as a result of the DNA shuffling of pNB esterase, DNA shuffling may be less efficient for discovery of new mutations compared to a controlled mutagenesis technique (a beneficial mutation can be masked in the background of recombined sequences). Utilizing more than two parents for recombination introduces a statistical preference for not incorporating mutations in progeny, and this has several consequences especially with respect to screening. Recombination favors the dilution of progeny containing the most mutations, which has the effect of exponentially increasing the number of progeny that must be screened in order to find the rarest ones. Because shuffling large numbers of parent sequences can yield many possible combinations, it may also be necessary to strictly limit the number of parent sequences in any given recombination experiment. We have described two alternative search strategies which reduce the required number of variants examined, at the cost of possibly missing intermediate beneficial combinations.

Finally, recombination requires a population of positive variants for efficient enzyme improve-

ment. If a population of positive variants must first be generated, sequential random mutagenesis may require less effort to produce sequences containing multiple mutations. Multiple positive variants are often generated, however, during a single cycle of random mutagenesis and screening. Recombination of these positives can provide substantial improvements at relatively little cost.

Materials and Methods

DNA shuffling

DNA shuffling was performed as described by Stemmer (1994b) with modifications. The 2 kb DNA fragment encoding the *B. subtilis* pNB esterase gene was amplified using PCR (forward primer 5′-CAATCTAGAGGGTATTAATAATG-3′ and reverse primer 5′-CGCGGGATCCCCGGGTACCGGGC-3′). The amplified DNA was purified by gel electrophoresis and extraction using Qiaex kit (Qiagen, Chatsworth, CA). A total quantity of ~10 μg DNA, either from a single parent (non-recombinatorial) or from a mixture of multiple parent sequences (recombinatorial), was digested with DNase I (0.0015 units/μl) at room temperature for 20 minutes in a 100 μl reaction. After ethanol precipitation, the digested DNA was electrophoresed as a smear in a 3% low melting temperature gel of NuSieve GTG Agarose (FMC Bio Products, Rockland, ME). DNA fragments in specified molecular size ranges were collected onto DE-81 filter paper disks (Whatman, Maidstone, England) by electrophoresis and eluted from the filter paper with 400 μl of 10 mM Tris/1 mM EDTA buffer (pH 8.0) containing 1 M NaCl. The DNA fragments were ethanol precipitated and redissolved to approximately 20 ng DNA/μl in 1 × Pwo DNA polymerase buffer (Boerhinger Mannheim, Indianapolis, IN) containing 2 mM $MgSO_4$ and 0.2 mM each of the four dNTPs. A 5 unit/μl Pwo DNA polymerase solution (Boehringer Mannheim) was diluted tenfold, and 0.5 μl was added to a 5 μl redissolved DNA reaction solution. Reassembly of DNA fragments was conducted by PCR, using the conditions 94°C for 40 seconds., then 70 cycles of 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 30 seconds, followed by a final extension step at 72°C for five minutes. A second 0.5 μl of Pwo polymerase was added at the annealing step of the 35th cycle. The reassembled DNA fragments were amplified in a conventional PCR (25 cycles) with the dilution of 1 μl reassembled DNA fragments in a 100 μl reaction. Once the success of the reassembly/amplification reactions was verified by gel electrophoresis, the reassembled product was purified with a Wizard PCR prep kit (Promega Corp., Madison, WI), digested with *Bam*HI and *Xba*I, concentrated by ethanol precipitation, and electrophoresed in an agarose gel. The 1.8 kb product was cut from the gel and the DNA extracted using a Qiaex kit. The final products were ligated with the vector generated by *Bam*HI-*Xba*I digestion of pNB106R (Zock *et al.*, 1994). This library was used to transform competent *E. coli* TG1 cells, as described (Moore & Arnold, 1996).

Screening a pNB esterase library

Screening was based on the assay described previously (Moore & Arnold, 1996), using the *p*-nitrophenyl ester of the loracarbef nucleus (LCN-pNP) as substrate. *E. coli* TG1 containing the plasmid library were grown on LB/tetracycline (20 μg/ml) plates. After 36 hours at 30°C single colonies were picked into 96-well plates containing 100 μl LB/tetracycline medium per well. These plates were shaken and incubated at 30°C for 12 hours to let the cells grow to saturation. Aliquots (20 μl) of the cultures were inoculated into a fresh plate containing 100 μl media per well; these were incubated at 40°C for ten hours with shaking to induce the expression of pNB esterase. Esterase activities were then measured by transferring 20 μl aliquots of the cell cultures into a fresh set of plates where they were mixed with 200 μl of 0.1 M Tris-HCl (pH 7.0) 25% DMF and 2 mM LCN-pNP. Reaction velocities were measured at 450 nm over ten minutes. (11 data points) in a ThermoMax microplate reader (Molecular Devices, Sunnyvale CA). Activities were normalized by the cell densities of the original wells measured at 600 nm to control for variations in cell quantities.

For each round of screening, the clones that showed the highest activities were re-streaked on LB/tetracycline agar plates, and single colonies derived from these plates (three to four colonies from each clone) were inoculated simultaneously into 96-well plates and tube cultures. The former were used to repeat the activity assay, as described above, and the latter were used for glycerol stock and plasmid preparation (Qiawell kit, Qiagen).

Assay of pNB esterase activity on LCN-pNB

A modified HPLC assay was used to determine enzyme activity towards the LCN-pNB (*p*-nitrobenzyl ester) substrate (Chen *et al.*, 1995). The bacterial cells were incubated at 30°C with shaking for 12 hours and then at 40°C for ten hours to induce expression of pNB esterase. Aliquots of cells (200 μl) were incubated with 300 μl reaction buffer for 30 minutes at room temperature. The final reaction mixtures contained 0.1 M Tris-HCl (pH 7.0) 25% DMF and 2 mM LCN-pNB. The reactions were stopped by addition of 500 μl acetonitrile and passed through a nylon syringe filter (Micron Separations, Inc., Westboro, MA) with a pore size of 0.45 μm. Aliquots of each sample (50 μl) were analyzed by HPLC on a 250 mm × 4.6 mm C18 reverse-phase column (Vydac, Hesperia, CA) at room temperature using a linear gradient starting with 50:50 of A:B (A = 5% methanol/95% 1 mM triethylamine, pH 2.5; B = 100% methanol) and ending with pure B in eight minutes (flow rate of 1 ml per minutes). Product and substrate were detected at 270 nm. The area of the *p*-nitrobenzyl alcohol product peak was calculated and subtracted from the area of the same peak from a sample containing *E. coli* without a pNB esterase gene. This controls for the small quantities of free product in the substrate preparation and any interference from bacterial contamination. This final area was used as a measure of total activity, which was normalized by cell density.

Acknowledgments

The authors thank Dr W. P. C. Stemmer (Maxygen) for many helpful discussions, Dr Steve Queener (Eli Lilly & Co.) for providing us the wild-type pNB esterase and the challenge to evolve it, and Ms Rebecca Little (Eli Lilly & Co.) for assistance with DNA sequencing. This

research is supported by the US Department of Energy's program in Biological and Chemical Technologies Research within the Office of Industrial Technologies, Energy Efficiency and Renewables. O. K. is supported by an NIH predoctoral training fellowship from the National Institute of General Medical Sciences, NRSA Award 1 T32 GM 08346-01.

References

Chen, K. & Arnold, F. (1993). Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. USA*, **90**, 5618–5622.

Chen, Y., Usui, S., Queener, S. W. & Yu, C. (1995). Purification and properties of *p*-nitrobenzyl esterase from *Bacillus subtilis*. *J. Ind. Micro.* **15**, 10–18.

Crameri, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. C. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotech.* **14**, 315–319.

Maynard Smith, J. (1988). The Evolution of Recombination. In *The Evolution Of Sex: An Examination Of Current Ideas*, pp. 106–125, Sinauer Associates, Inc, Sunderland, Mass.

Moore, J. C. & Arnold, F. H. (1996). Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nature Biotech.* **14**, 458–467.

Moore, J. C. & Arnold, F. H. (1997). Optimization of industrial enzymes by directed evolution. *Advan. Biochem. Eng.* **58**, 1–14.

Müller, H. J. (1932). Some genetic aspects of sex. *Amer. Nature*, **66**, 118–138.

Shafikhani, S., Siegel, R. A., Ferrari, E. & Schellenberger, V. (1997). Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. *Biotechniques*, in the press.

Stemmer, W. P. C. (1994a). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature*, **370**, 389–391.

Stemmer, W. P. C. (1994b). DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl Acad. Sci. USA*, **91**, 10747–10751.

Wells, J. A. (1990). Additivity of mutational effects in proteins. *Biochemistry*, **29**, 8509–8517.

You, L. & Arnold, F. H. (1996). Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng.* **9**, 77–83.

Zhao, H. & Arnold, F. H. (1997a). Optimization of DNA shuffling for high fidelity recombination. *Nucl. Acids Res.* **25**, 1307–1308.

Zhao, H. & Arnold, F. H. (1997b). Functional and non-functional mutations distinguished by random recombination of homologous genes. *Proc. Natl Acad. Sci. USA*, **94**, 7997–8000.

Zock, J., Cantwell, C., Swartling, J., Hodges, R., Pohl, T., Sutton, K., Rosteck, P., Jr, McGilvray, D. & Queener, S. (1994). The *Bacillus subtilis* pnbA gene encoding *p*-nitrobenzyl esterase: cloning, sequence and high-level expression in *Escherichia coli*. *Gene*, **151**, 37–43.

Edited by J. Wells

(Received 10 March 1997; received in revised form 18 June 1997; accepted 9 July 1997)

Current Opinion in Biotechnology

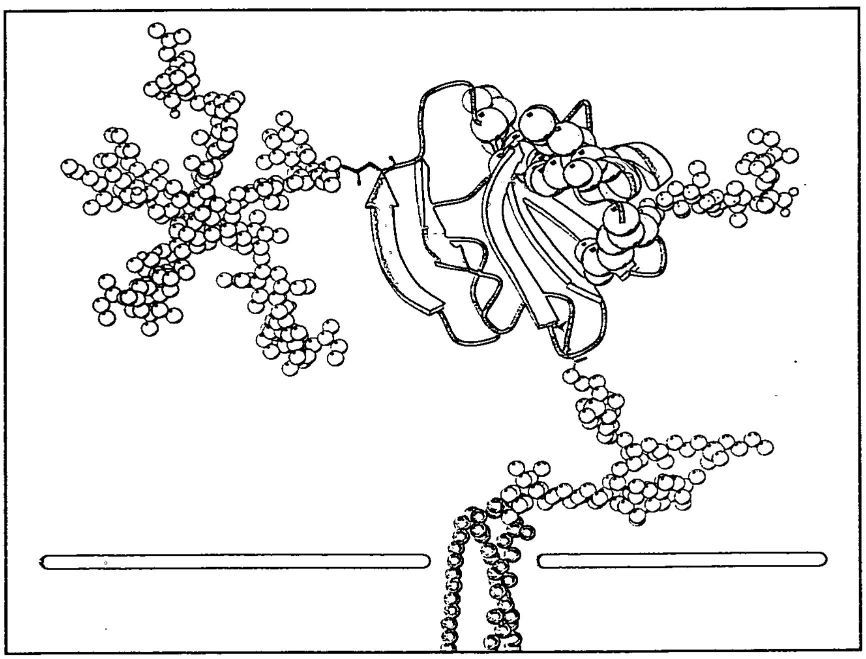
Julian Davies & Martin Rosenberg, Editors

Vol 8 No 4 August 1997

Protein engineering

Edited by Birgit A Helm and Eduardo A Padlan

Web alert



Current Biology Ltd

Applications of DNA shuffling to pharmaceuticals and vaccines

Phillip A Patten*, Russell J Howard† and Willem PC Stemmer‡

DNA shuffling is a practical process for directed molecular evolution which uses recombination to dramatically accelerate the rate at which one can evolve genes. Single and multigene traits that require many mutations for improved phenotypes can be evolved rapidly. DNA shuffling technology has been significantly enhanced in the past year, extending its range of applications to small molecule pharmaceuticals, pharmaceutical proteins, gene therapy vehicles and transgenes, vaccines and evolved viruses for vaccines, and laboratory animal models.

Addresses
Maxygen, Inc., 3410 Central Expressway, Santa Clara, CA 95051, USA
*e-mail: phil_patten@MAXYGEN.com
†e-mail: russell_howard@MAXYGEN.com
‡e-mail: pim_stemmer@MAXYGEN.com

Current Opinion in Biotechnology 1997, 8:724–733

http://biomednet.com/elecref/0958166900800724

Abbreviations

IFN	interferon
kb	kilobases
MLV	murine leukemia virus

Introduction

The sequence-design processes used by nature have yielded results far superior to those obtained so far by rational approaches to the design of biological structures and systems. The rational design of proteins, for example, is performed by computer modeling of individual changes, followed by construction of the corresponding DNA and expression of the recombinant protein for testing and evaluation. This approach seeks to design proteins for specific tasks by drawing correlations between amino acid sequence and specific shapes in an attempt to understand a complex system through its topography. At its root, our limited ability to rationally engineer complex biological systems stems from three main factors: limited structure/function knowledge of most proteins and of their complex interactions with expression machinery and other molecules within the cell; the computationally intensive and approximate nature of the modelling required; and the fact that the number of relevant mutants for which one would wish to make predictions in order to optimize a given function ranges from millions to numbers of astronomical proportions [1].

In contrast to rational design procedures, nature employs mutation, selection and recombination to evolve highly adapted individuals from the effectively infinite possibilities encoded implicitly in the genome. Recent technological advances have demonstrated that it is now possible to mimic these natural evolutionary processes. Benchtop *in vitro* evolution techniques are used to construct libraries as large as 10^{15} molecules [2–5]. Researchers can mimic natural evolution by searching these libraries by, for example, affinity panning of phage-displayed or RNA ligands against pharmaceutical targets, for the best candidates for a specific task. Repeated rounds of selection and amplification of candidates has already produced improved enzymes and novel molecules capable of binding their targets with higher affinity than their natural counterparts. Unlike natural selection, in which multiple environmental forces select organisms with genomes that allow them to meet a variety of challenges, *in vitro* evolution exerts focused selection pressure on organisms in isolation, enabling the rapid development of variants with highly specialized traits.

Despite the enormous potential of these techniques, determination of the best strategy to exploit this diversity has been the topic of much debate. The most popular methods of creating combinatorial libraries are strategies that seek to evolve sequences that have individual point mutations or blocks of oligonucleotide encoded mutations. At present, most researchers use either repeated cycles of 'error-prone PCR' [6,7] or repeated oligonucleotide directed mutagenesis [8] to create these 'point mutation' libraries. Error-prone PCR employs a low fidelity replication step to introduce random point mutations at each round of amplification [6]. This method has the advantage of simplicity and ease of use. The power of these methods is limited, however, by one's ability to identify critical regions for mutagenesis and because, generally, only small regions of the genome can be mutagenized to saturation and be exhaustively sampled in screens or selections, due to limited library sizes relative to the size of the sequence spaces defined by exhaustive random searches [9].

Iterative cassette or point mutagenesis can overcome some of these limitations; however, as discussed below, DNA shuffling profoundly accelerates the process. In this review we summarize recent advances which extend the range of application of this technique to small molecule pharmaceuticals, pharmaceutical proteins, gene therapy vehicles and transgenes, vaccines, and evolved viruses for laboratory animal models of disease.

Recombination

A key limitation of random point mutagenesis and random cassette mutagenesis strategies can be traced to the fact that they introduce random 'noise' into the gene population at every cycle, and hence improvements are limited to small steps. If the noise level is too high relative to the library size and the selection stringency, the

message will gradually become riddled with deleterious mutations. This is analogous to the phenomenon of Muller's Ratchet [10] from population biology in which, in the absence of sexual recombination, deleterious mutations build up in a population over time. From decades of plant and animal breeding, classical breeders have learned to use recombination to rapidly evolve improved sequences for a specific task. Sexual replication in combination with directed selection can produce substantial improvements in a highly diverse genome within just a few cycles. The widespread prevalence in nature of sexual versus asexual reproduction has been the topic of much debate [11]. A gene favouring parthogenic rather than sexual reproduction would quickly take over in a population unless counterveiled by selection, the so-called twofold cost of sex. One hypothesized major force for the maintenance of sexual reproduction is that recombination of natural diversity allows for populations to rapidly evolve in order to adapt to changing physical environments or pathogens by combining beneficial mutations into single, more fit individuals and deleterious mutations into less fit individuals that are then selected out of the gene pool [11].

The technique of DNA shuffling comes the closest of any laboratory technique to mimicking natural recombination by allowing the *in vitro* homologous recombination of DNA [12,13]. This technique not only recombines DNA fragments, but also introduces point mutations at a very low, controlled rate [14], thus, combining recombination, point mutation, and selection techniques to create a general, parallel algorithm for evolving improved genes. This parallel search strategy is analogous to that used by massively parallel supercomputers [9], such as those used for climate modelling or to factor very large prime numbers. Figures 1 and 2 summarize the practical and theoretical advantages of DNA shuffling relative to existing recursive mutagenesis methods, such as error-prone PCR or recursive oligonucleotide directed mutagenesis. Recent progress with DNA shuffling has clearly demonstrated the utility of recombination for accelerating molecular evolution through simultaneously permuting both single mutations and large sequence blocks (Table 1). DNA shuffling combined with focused selection pressure in the laboratory will allow one to rapidly evolve genes for a wide variety of industrial applications: the optimization of enzymes, such as proteases, lipases, amylases and cellulases; the development of metabolic pathways specialized to synthesize large amounts of specialty chemicals, antibiotics, or pharmaceutical proteins; organisms designed for bioremediation; and plasmids or viruses for novel vaccines and gene therapy applications (Table 2). These emerging applications are discussed below.

Family shuffling dramatically increases the rate of st pwise evoluti n

Forced hybridization between species, such as was performed in the cross-breeding of the plum and apricot to yield the plumcot, was recognized last century by plant breeders as a highly effective method for generating novel, functional varieties with phenotypes differing dramatically from either parent [15]. The utility of this strategy has now been demonstrated at the level of a single gene. To evaluate whether recombining natural diversity accelerates the evolution process, the efficiency of obtaining a new substrate specificity from four homologous enzyme genes evolved separately was compared with that from a recombined pool of the four genes. The essential goal was to compare the rate of evolution of a single gene that is subjected to random mutation and selection to the rate of evolution of a library of genes created by recombining existing, functional genetic diversity present in a family of homologous genes. Since all of the recombinants are created from diversity that has proven functional in the context of its parental gene, the hope is that such libraries would be of much higher quality than random libraries. The results affirm this view. One cycle of single gene shuffling yielded eightfold improvements from each of the four separately evolved genes, versus a 270–540-fold improvement from the four genes shuffled together ([16••]; Figure 3a). This represents an approximately 50-fold increase in the rate of improvement per cycle. The best clone contained eight segments from three of the four genes as well as 33 amino acid point mutations. It is worth emphasizing that this evolved improvement relative to the initial gene pool was obtained in a single cycle of gene shuffling, rather than requiring many recursive cycles.

Thus, in contrast to classical breeding techniques, DNA shuffling allows one to readily recombine DNA derived from 'separate' species or genera. This results in a much more sparse sampling of sequence space (Figure 3b), in which the average similarity between library members is much lower than with other strategies. Sparse sampling yields mutants that, after a single cycle, are far more divergent from the parental genes than is possible with single gene shuffling or point mutation strategies. This recent experiment demonstrates that cross-species recombination is a remarkable accelerant of molecular evolution. Family shuffling will be widely applicable to the commercially important problems discussed below.

Protein pharmaceuticals

Recombinant pharmaceutical proteins form a multi-billion dollar sector of the pharmaceutical industry. This industry relies principally on cloning existing genes encoding cytokines, growth factors, and enzymes that are the products of millions of years of evolution. Many of these products have side effects that cause serious complications which limit or preclude their clinical use. Selective breeding using DNA shuffling provides a technology for rapidly improving pharmaceutical proteins through selective breeding for enhancement of desirable biological activities, while eliminating or reducing undesired activities (Table 2).

Table 1

Genes and operons evolved by DNA shuffling.

System	Improvement	Size	Cycles	Mutations	Comments	Reference
TEM-1 β-lactamase	Enzyme activity 32,000-fold	1 kb	3 + 2†	6 aa	Selection MIC 0.02 μg/ml to 640 μg/ml In comparison, three cycles of mutagenic PCR and selection=16-fold	[12]
β-lactamase family	Enzyme activity 270–540-fold	1 kb	1	Chimerics	Selection 500-fold jump in fitness in one round of shuffling and screening 50,000 colonies	[16]
β-galactosidase	Fucosidase activity 66-fold Substrate specificity 1,000-fold	4 kb	7	6 aa 13 bp	Screen 10,000 colonies per round, best 20–40 chosen per cycle 66-fold fucosidase activity: 2–3-fold increase expression plus 20-fold increase activity 11/13 base mutations in coding sequence	[30••]
Green fluorescence protein	Protein folding 45-fold (*E. coli* and mammalian cells)	0.8 kb	3	3 aa 6 bp	Screen 10,000 colonies per round, best 20–40 chosen per cycle Protein folding improvement	[31]
Antibody (scFv)	Avidity >400-fold	0.8 kb	8 + 2†	34 aa	Phage panning Naive human antibody library Stop codon modified to suppressible stop codon	[32]
Antibody (scFv)	Expression level 100-fold	0.8 kb		5	Phage panning Murine hybridoma library	[32]
Arsenate operon	Arsenate resistance 40-fold	2.3 kb	3	3 aa 13 bp	Selection Three genes; 5.5 kb plasmid shuffled 10,000 colonies per cycle All three mutations in arsB efflux pump, arsC was predicted to be rate limiting Arsenate reduction activity up 5-10-fold Growth in 0.5 M arsenate Plasmid integrated in round 3	[19]
Atrazine degradation	Atrazine degradation 80-fold	5.6 kb	4	8 aa‡	Screen Two genes: 8.5 kb plasmid shuffled Atrazine chlorohydrolase activity increased Wild-type inactive versus terbutylazine, evolved enz active Screen: insoluble atrazine to soluble product gives zone of clearing	
Alkyl transferase	DNA repair 10-fold	0.5 kb	6	7 aa	Selection Suicide enzyme: limited potential for improvement	FC Christians, G Dawes, WPC Stemmer, unpublished data
Benzyl esterase	Antibiotic deprotection 150-fold	1.5 kb	2	8 aa	Screen Four rounds of PCR mutagenesis and screening for improved variants prior to shuffling	[20]
tRNA synthetase	Charging of engineered tRNA 180-fold	2.0 kb	7	nd	Selection	[22]

†Additional backcross cycles. ‡Rare mutation: 3 NA insertions in 5 codons restored frame, inserted 1 aa with 5 other aa changed atz gene lost 2 kb in final mutant, no loss of function. aa, amino acid changes; bp, base pair changes; nd, not determined.

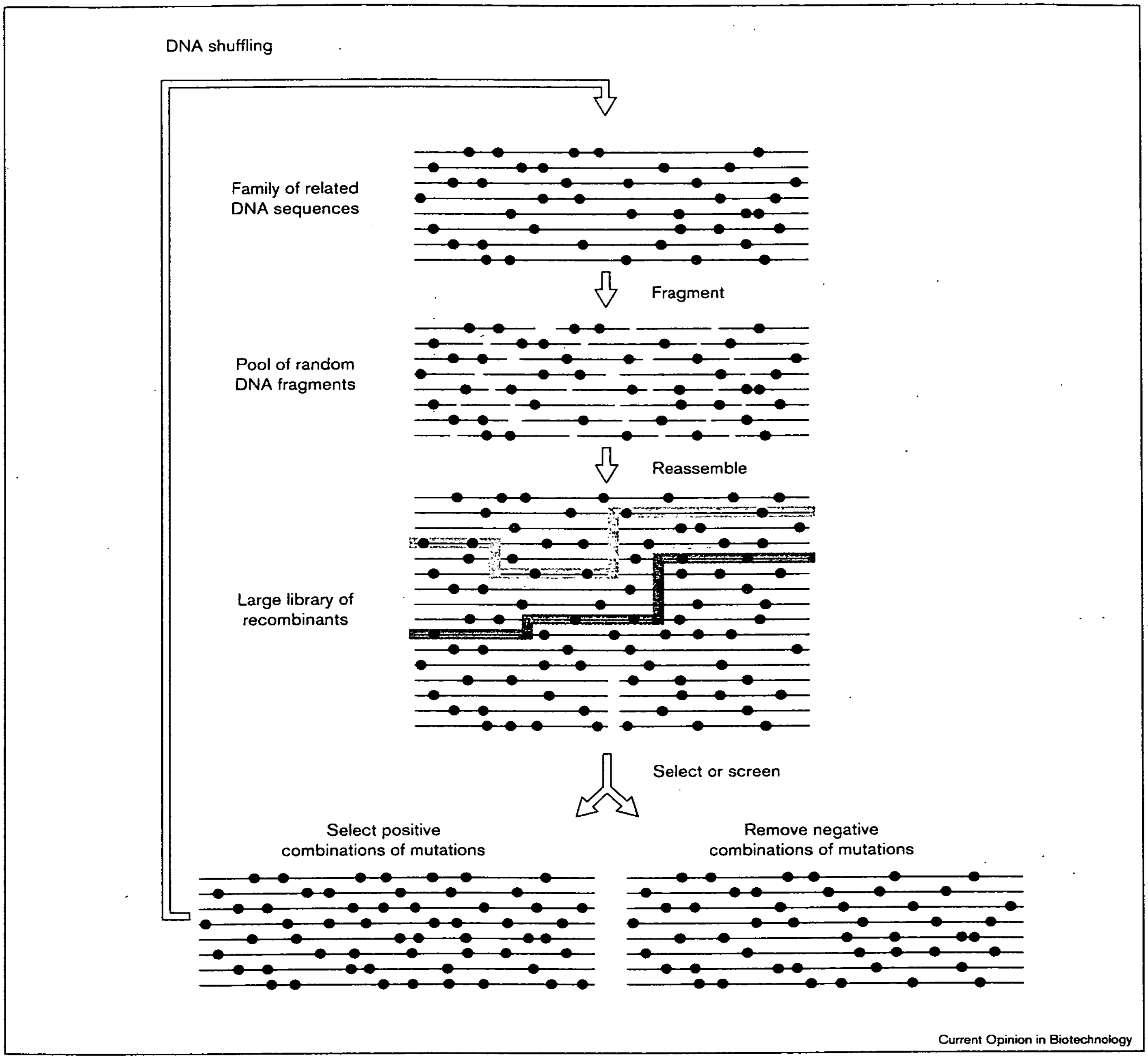
We have recently applied DNA shuffling to members of the human α-IFN gene family (85–97% pairwise amino acid identity). Greater than 10^{26} distinct recombinants can be generated from the natural diversity in this gene family (Figure 4a). While no foreseeable library technology will allow an 'exhaustive' sampling of this sequence space, DNA shuffling technology provides a design algorithm with which to selectively breed for IFNs with increased potency relative to naturally occurring IFNs. Typical chimeras produced by DNA shuffling are shown in Figure 4b. Generic high-throughput methods for α-IFN expression and biological assay as fusion proteins on phage have been developed and used for rapid parallel analysis of recombinant IFNs. Phage-displayed recombinants with improved potency on human and murine cells have been obtained (Figure 4).

This generic approach can be applied to improve many pharmaceutical proteins. Proteins with novel activities have been created by directed chimerization of modules from existing pharmaceutical proteins [17,18], a strategy that is likely to be particularly effective with cytokines, which typically act by dimerizing two or more receptor components. As with the rapid evolution of moxolactamase activity by shuffling a cephalosporinase gene family [16••], many new pharmaceutical activities may be discovered through breeding large libraries of chimeric pharmaceutical proteins. Selective breeding using DNA shuffling will allow rapid evolution of pharmaceutical proteins with potent activities from such recombinants, which initially have low levels of the desired activity. Backcrossing of these evolved variants with the wild type genes will allow one to remove functionally neutral changes, thus reducing the immunogenicity of the evolved proteins.

Small molecule pharmaceuticals and industrial enzymes

Microorganisms are widely used for the production of pharmaceutical molecules, such as antibiotics, antifungals, anticancers and immunosuppressives. In many cases, the genes encoding the relevant biosynthetic enzymes are known, often occurring in operons or gene clusters. Rational engineering of these biosynthetic pathways to improve yield or generate analogs is difficult because,

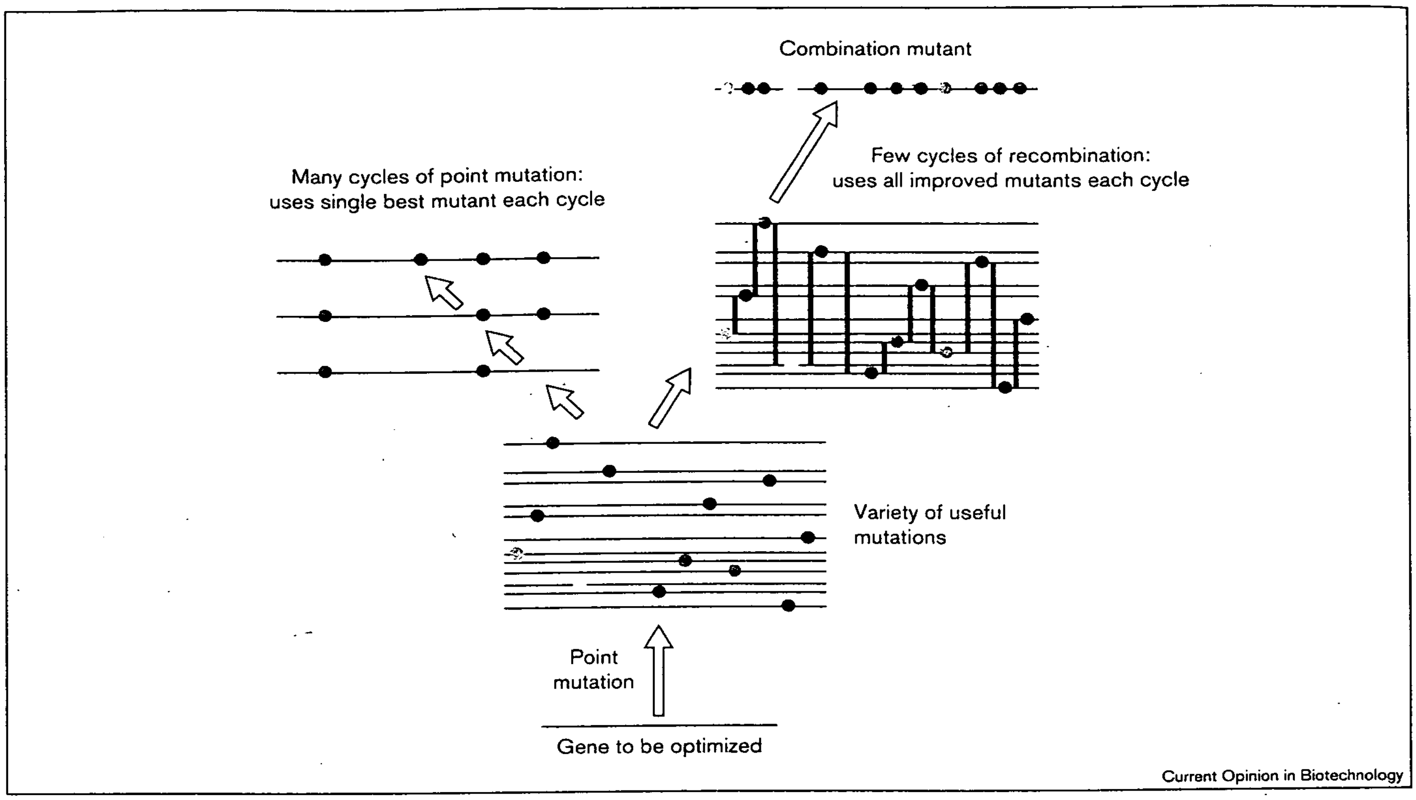
Figure 1



DNA shuffling methodology. The first step of this method is to randomly fragment a population of related genes using DNAse I. This produces fragments of various lengths that, after denaturation, hybridize to form an equal mixture of 5′ and 3′ overhangs. Using PCR techniques, the 5′ overhang fragments can be extended by Taq DNA polymerase – leaving the 3′ overhang fragments unaffected. As a consequence of this extension, the average fragment length increases during each cycle. Recombination occurs when a fragment derived from one template primes a template with a different sequence [13]. Green dots represent beneficial mutations and red dots represent deleterious mutations. The coloured bars indicate recombinations of portions of three parents into recombinant progeny.

in addition to the difficulties of protein engineering, the determination of rate limiting steps in the pathway is laborious and uncertain. DNA shuffling is well suited to the optimization of such pathways because the entire pathway can be treated as the unit for evolution, with no requirements for knowledge of the rate limiting steps or for detailed structure/function analysis of the proteins. Pathways for the detoxification of atrazine (J Minshull, personal communication) and arsenate [19•] have been improved using DNA shuffling (Table 1). Importantly, in these examples no *a priori* knowledge was needed to yield significant improvement. A benzylesterase used

Figure 2



The advantages of DNA shuffling as a sequence design algorithm for evolving complex new gene functions are shown schematically. The coloured dots represent beneficial mutations. The vertical direction represents a generalized measure of fitness (i.e. kcat/Km for an enzyme) with the fitest genes being at the top. Because the frequency of beneficial mutations is generally low relative to deleterious mutation, only single beneficial mutations are generally added in each cycle of random mutagenesis and screening or selection. Hence, procedures that use iterative point mutations must build up beneficial mutations one at a time through many rounds of selection, generally with only the best mutant from any given cycle being pursued. In contrast, DNA shuffling allows one to directly recombine all beneficial mutations from any given round into multi-step mutants with dramatically improved phenotypes.

industrially for deprotecting a precursor of the antibiotic loracarbef has been improved using DNA shuffling ([20]; Table 1). Recombination was shown to be superior to sequential error-prone point mutation for the evolution of this activity [21•]. DNA shuffling has also recently been used to modify the specificity of a tRNA charging enzyme [22], with the ultimate goal of evolving tRNA synthetases that can specifically charge tRNA's with unnatural amino acids incorporated at specific sites.

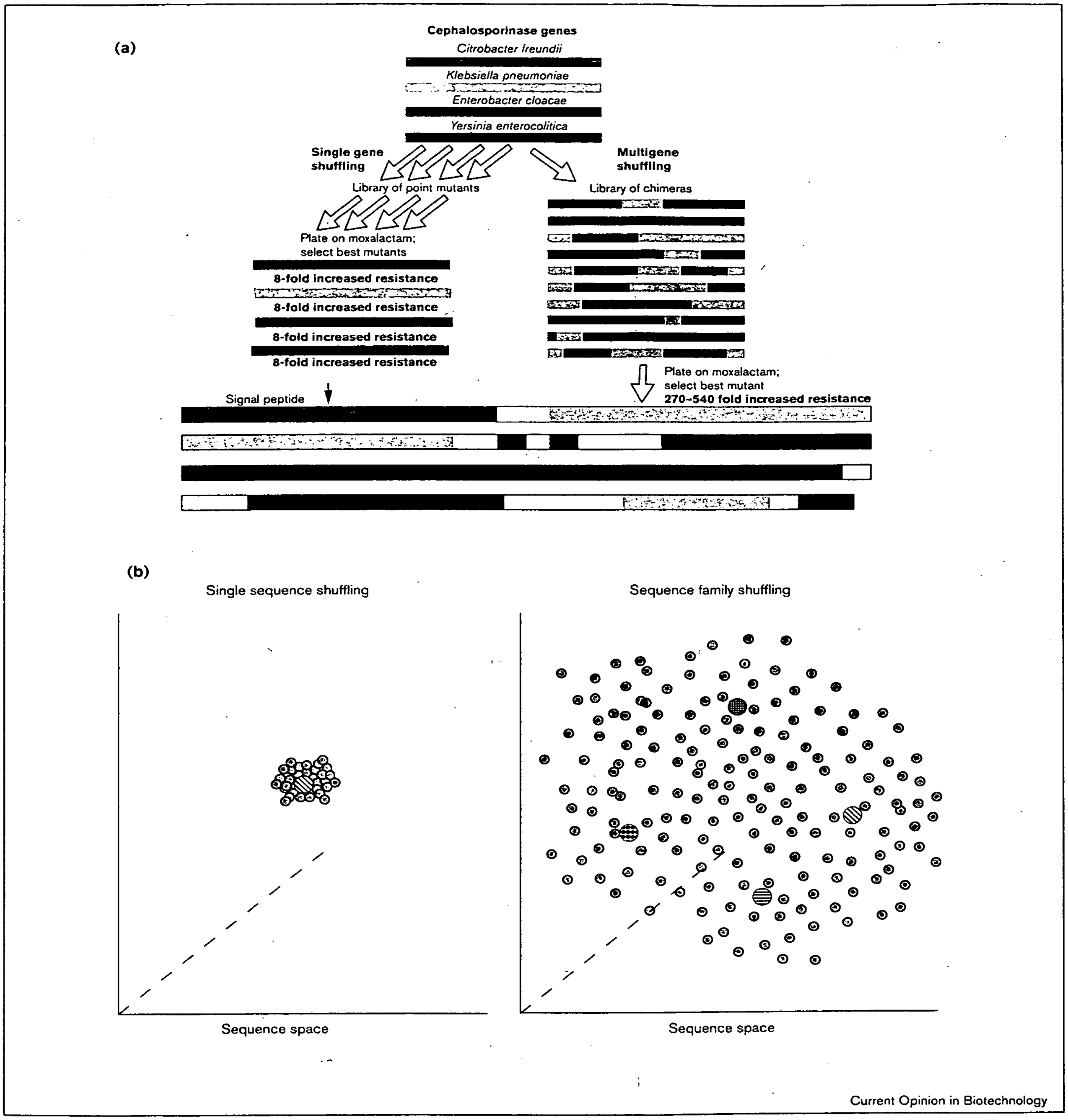
Evolved enzymes will find wide application in the replacement of multi-step chemical reactions required for manufacture of drugs or their precursors by an enzymatic conversion. Most naturally occurring enzymes capable of such valuable conversions require significant modification in activity, specificity, or expression level before they are suitable for large scale drug manufacture. DNA shuffling provides an important tool for the optimization of such enzymatic conversions.

Evolved viruses for pharmaceutical applications

The full length genomes of many viruses are in the range of 5–15 kilobases (kb), a size range that can be readily handled by current DNA shuffling methods (Table 1). Our ability to clone and sequence the wealth of natural viral isolates far outstrips our molecular understanding and our ability to rationally manipulate them. Three wild-type strains of human papilloma virus have been successfully shuffled (D Apt, personal communication). The biological properties of this library of recombinants are currently being investigated. This approach has potential for the evolution of human papilloma virus to overcome the blocks to growth in transformed fibroblasts, and thus be able to grow in readily manipulated tissue culture systems for drug screening.

Adenovirus is widely used as a gene therapy vehicle. Over 100 naturally occurring serotypes with differing

Figure 3



(a) The strategy and results from shuffling four homologous cephalosporinase genes are shown schematically. Single gene shuffling resulted in eightfold increased resistance to the antibiotic moxolactam, whereas shuffling the gene family gave a 270–540-fold increase in resistance in a single step [16]. **(b)** Evolution starting from a single gene is schematically contrasted with evolution based on shuffling a homologous gene family. The axes denote a generalized sequence space. Shaded dots indicate particular sequences present in a given library. Hatched dots represent sequences that are more fit than the best parental molecule. Greater distance indicates greater sequence divergence. Family shuffling results in a relatively 'sparse' sampling of sequence space with relatively few individuals that are highly similar to the parental molecules and many individuals that are very divergent from the parents.

Table 2

Development of novel human therapeutics through molecular breeding technology.

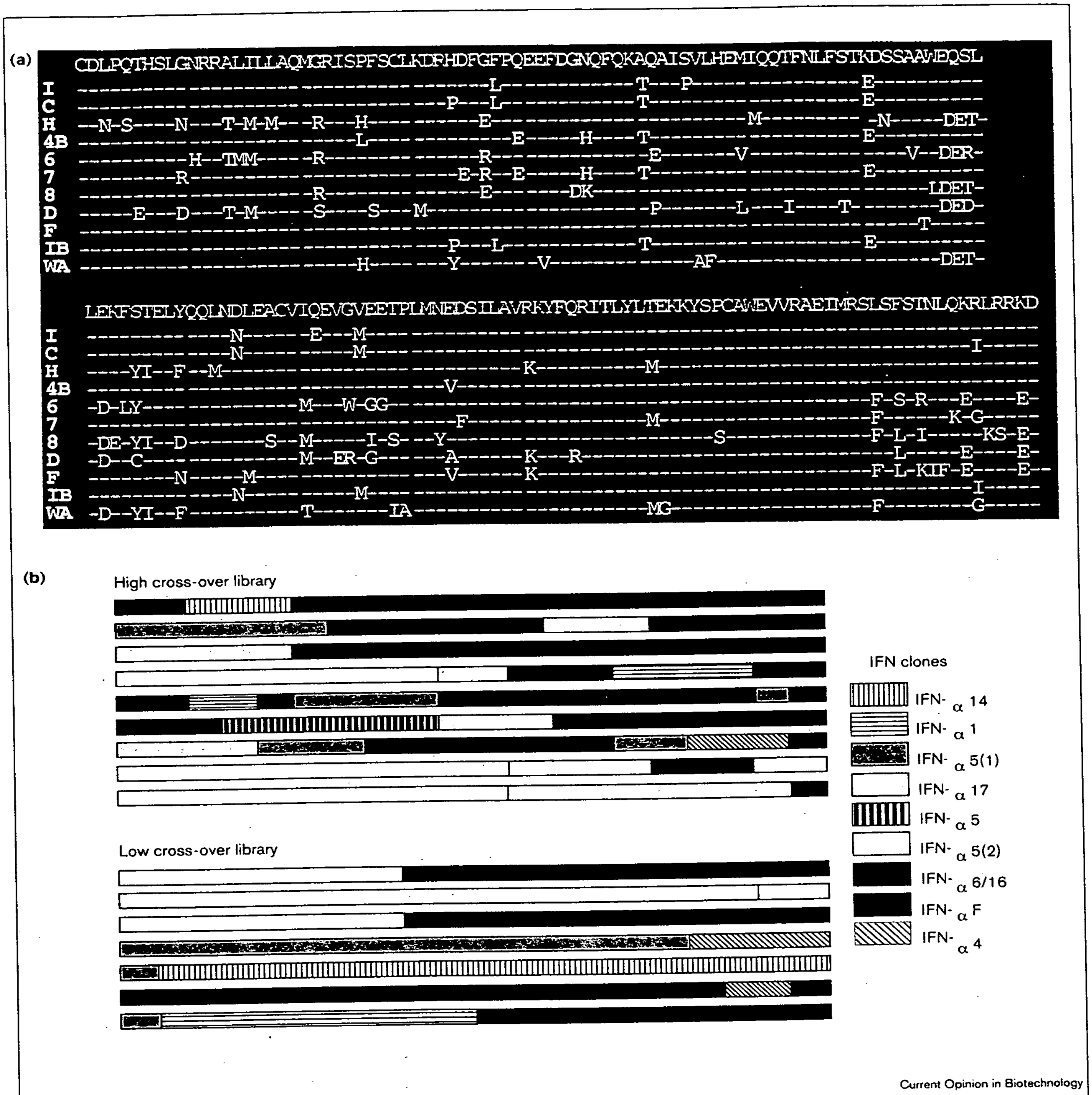
Therapeutic area	Size of shuffled DNA	Novel properties for screening or selection
Protein pharmaceuticals	1–5 kb	Receptor selectivity Improve agonist/antagonist activity Novel agonist/antagonist activity Toxicity Optimal expression in desired host – shuffle the expression vector and/or protein gene Plasma half-life/protein and cell binding
Small molecule pharmaceuticals	1–5 kb	Manufacturing enzymes: desired properties for enzymes that replace chemical synthetic steps in drug manufacture (pH and temperature optimization, solvent tolerance, etc.)
	1–20 kb	Natural products: generation of drug analogs/new structures by shuffling entire pathways – screen for pharmaceutical activity Strain improvement
Gene therapy	1–20 kb	Vectors Transgenes
DNA vaccines	1–20 kb	Plasmid properties of promoter strength, tissue tropism and cell entry, plasmid stability Antigen shuffling to optimize immune response Shuffling to improve immunostimulatory sequences Qualitative nature of immune response (shuffling cytokine gene cassettes)
Recombinant protein vaccines	1–20 kb	Optimal expression in desired host – shuffle the expression vector and/or the protein gene while selecting for retention of recognition by protective mAbs Shuffling to generate family of immunogens for antigenically variant targets
Viral vaccines	1–20 kb	Enable viral growth/optimize yield in desired cell for manufacturing Select for retention of desired properties when viral families are shuffled – neutralization assay, antibody effects Select for attenuation (growth rate, temperature, retention of some properties) Select for optimal expression of recombinant antigen(s) genes in viral vaccine vectors
Evolved viruses	1–20 kb	Evolve to grow in animal models such as mice Evolve viruses to grow in easily manipulated tissue culture systems

tropisms are known. The fiber and penton genes of adenovirus are the major determinants of tissue tropism, as they are responsible for cell adhesion. The pentons of adenoviruses interact with cellular integrins [23] and the fibers with cellular receptors, one of which has been newly identified [24]. Evolved adenoviral variants which could be selectively targeted to particular cell types would be of great utility for gene therapy and vaccine delivery vectors. The penton and fiber genes of various adenovirus serotypes have been shuffled (S Liu, personal communication) and this approach may allow one to evolve mutant adenoviral vectors with desired cell and tissue tropism.

Murine leukemia virus (MLV) is a retroviral vector that has received much attention as a gene therapy vehicle. As with adenovirus, there are <100 naturally occurring strains, whereas only a single MLV strain is being developed for gene therapy applications (Moloney MLV). The envelopes of 15 MLV strains have been shuffled and viruses are being selected for improved tropism, titer, stability and gene expression properties (N Soong, personal communication).

HIV-1 poses a major threat to human health which is increasing because of the growing viral load worldwide, the high rate of evolution of this pathogen, and the concomitant evolution of associated opportunistic pathogens in HIV-1 infected individuals. There is currently no practical animal model in which to test the multitude of antiviral drugs or vaccine strategies [25]. Work is beginning on the genetically engineered animal models to support replication of HIV-1 (D Littman, M Goldsmith, personal communication), but no replication has yet been observed in these hosts. It is clear from viral phylogenetic trees that lentiviruses can evolve the ability to grow in new species and it is clear that recombination plays a major role in the natural high rate of evolution of lentiviruses [20]. DNA shuffling provides a powerful new tool with which to accelerate the adaption of viruses to grow in laboratory animal hosts. Recombination is believed to be of great importance for the naturally high rate of evolution of retroviruses [26]. Laboratory animals have been engineered with the human HIV-1 receptor and co-receptor genes (D Littman, personal communication). DNA shuffling is being used to recombine entire genomes and individual genes of natural HIV-1 isolates to accelerate the adaptation of HIV-1 to grow in these engineered animals (P Patten and N Landau, unpublished data). We anticipate that the adaptation of HIV-1 to replicate in a laboratory animal will open up many fertile avenues for drug and vaccine discovery on this important human pathogen. Shuffling of natural diversity to create large libraries of chimeric viruses is a general approach that can be applied to other viruses, such as the hepatitis B

Figure 4



(a) Human α-interferon diversity. The sequences of eleven natural human α-IFN sequences are shown [29]. Consensus α-IFN is given at the top. Dashes indicate identity to consensus. The number of distinct recombinants that can be generated by shuffling these eleven genes is 3×10^{26}. This number is calculated by multiplying the number of different amino acids observed at each polymorphic site ($2^{56}\times3^{15}\times4^{4}=3\times10^{26}$). **(b)** Sixteen representative recombinant α-IFNs derived by shuffling eight natural human α-IFN genes are shown schematically (P Patten, unpublished data). The high crossover and low crossover libraries were generated by shuffling 20–50 bp or 50–100 bp fragments, respectively. The DNA shuffling was done essentially as described in [16]. The shuffled IFNs were expressed as fusions to gene III on bacteriophage M13 and screened for antiproliferative activity using Daudi cells. Phage displayed αIFN–MAX4 (high crossover library, fourth from the top) is 40-fold more active in a Daudi antiproliferation assay than IFN2α and twofold more active than consensus 1 IFNs displayed on phage.

virus and hepatitis C virus, for the purpose of producing domesticated forms of these viruses (i.e., easier to handle in the laboratory) for vaccine development and drug screening on variants that can readily be grown in tissue culture systems.

Vaccines

Development of effective vaccine technologies has stimulated renewed government emphasis and interest from pharmaceutical researchers. DNA vaccines are particularly attractive because of their relatively low cost and the feasibility of rapid generation of diverse variant vaccines containing evolved promoters, immunostimulatory sequences, cytokines, etc., for comparison testing [27]. For example, assays for selection of DNA vaccines with improved promoter activity, immunostimulatory sequences, enhanced expression levels or cell tropism can be developed to produce second generation DNA vaccine plasmids. Viral vaccine vectors can be enhanced by DNA shuffling to give desired properties of tropism, stability and expression level. The promise of this new technology notwithstanding, existing human vaccines rely on live or killed whole organisms or components purified from whole organisms. We envision opportunities for DNA shuffling as a tool for increasing the efficiency and success rate of the development of novel whole organism, viral, bacterial and recombinant protein vaccines. Pathogenic viruses can be subjected to DNA shuffling followed by selection for desired attenuation properties while retaining the immunogenicity required for a vaccine. Viruses that would serve as excellent vaccine vectors or as vaccines but which cannot be manufactured to sufficient titer in manufacturing cell lines, can be shuffled and selected for improved titre to create new commercial opportunities. Recombinant proteins that are known to be excellent vaccine immunogens but which cannot be manufactured in appropriate yield or in suitable host systems can be shuffled and screened to solve such expression problems, while co-selecting for retention of necessary epitopes. These and other valuable opportunities for application of DNA shuffling technologies in vaccinology are summarized in Table 2.

The impact of genomics

The rapid rate of increase in the availability of known gene sequences and our ability to manipulate them in cloned form greatly exceeds our understanding of these genes and our ability to engineer these sequences based on rational models. Sequence information from informatics databases can readily be converted into functional DNA clones given tools such as PCR, synthetic DNA and methods for the rapid assembly of genes from synthetic DNA [28].

Shuffling of natural diversity to explore the sequence space defined by shuffled homologues is a demonstrably powerful strategy for accelerated evolution of biological molecules with novel activities ([16]; Figure 3). We expect the dramatic success seen with family shuffling of cephalosporinase genes to be repeated in many other systems in which homologous genes are recombined. The explosion of DNA sequences available on the internet provides a rich, diverse and rapidly expanding supply of molecular breeding stock. DNA shuffling is a general and natural algorithm for functionally exploiting this natural diversity, with minimal requirements for *a priori* genetic or biochemical characterization to guide this exploitation.

Conclusions and future directions

In order to unlock desired biologically active sequences from the potential diversity present in an organism's gene pool, it is of great importance to understand which evolution algorithms are the most effective. Point mutation techniques, such as error-prone PCR and repeated oligonucleotide directed mutagenesis, search sequence space by creating libraries of randomly mutated molecules. In contrast, DNA shuffling exchanges large functional blocks of sequence containing previously selected mutations to search for the best candidate molecules, thus mimicking and accelerating the process of sexual recombination. We expect that, just as recombination has played a major role in the evolution of life, DNA shuffling will play a central role in the development of applied molecular evolution technologies and will prove indispensable for bringing existing biological diversity into the service of human health care, and agricultural and industrial chemical needs.

Acknowledgements

We thank Doris Apt, Scott Liu, Nai Wei Soong, Juha Punnonen and Jeremy Minshull for helpful discussions and for sharing unpublished results.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Kauffman SA: *The Origins of Order*. New York: Oxford University Press; 1993.
2. Joyce GF: **Directed molecular evolution.** *Sci Am* 1992, **267**:90-97.
3. Bartell DP, Szostak JW: **Isolation of new ribozymes from a large pool of random sequences.** *Science* 1993, **261**:1411-1418.
4. Tuerk C, Gold L: **Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.** *Science* 1990, **249**:505-510.
5. Breaker R, Joyce GF: **Emergence of a replicating species from an in vitro RNA evolution reaction.** *Proc Natl Acad Sci USA* 1994, **91**:6093-6097.
6. Caldwell RC, Joyce GF: **Randomization of genes by PCR mutagenesis.** *PCR Methods Applic* 1992, **2**:28-33.
7. You L, Arnold FH: **Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide.** *Protein Eng* 1994, **9**:77-83.
8. Reidhaar-Olson J, Bowie J, Breyer RM, Hu JC, Knight KL, Lim WA, Mossing MC, Parsell DA, Shoemaker KR, Sauer RT: **Random mutagenesis of protein sequences using oligonucleotide cassettes.** *Methods Enzymol* 1991, **208**:564-586.
9. Stemmer WPC: **Searching sequence space using recombination to search more efficiently and thoroughly**

instead of making bigger combinatorial libraries. *Bio-Technology* 1995, 13:549-555.

10. Muller HJ: **The relation of recombination to mutational advance.** *Mutat Res* 1964, 1:2-9.

11. Maynard Smith J: **The evolution of sex.** In *The Evolution of Sex.* Edited by Bellig R, Stevens G. San Francisco: Barper and Row; 1988:3-20.

12. Stemmer WPC: **Rapid evolution of a protein *in vitro* by DNA shuffling.** *Nature* 1994, **370**:389-391.

13. Stemmer WPC: **DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution.** *Proc Natl Acad Sci USA* 1994, **91**:10747-10751.

14. Zhao H, Arnold FH: **Optimization of DNA shuffling for high fidelity recombination.** *Nucleic Acids Res* 1997, 25:1307-1308.

15. Burbank L: **Short-cuts into the centuries to come: better plants secured by hurrying evolution.** In *Luther Burbank His Methods and Discoveries: Their Practical Application*, vol 1. Edited by Whitson J, Williams RJHS. New York: Luther Burbank Press; 1914:176-210.

16. •• Crameri AC, Raillard S, Stemmer WPC: **DNA shuffling of a family of genes from diverse species accelerates directed evolution.** *Nature* 1997, in press.

This paper describes the use of homologous gene family shuffling to obtain large jumps in fitness in a single cycle of DNA shuffling. This is the first paper to show the utility of family shuffling for rapid evolution of a gene and it is a technique which is expected to open up approaches to many new problems by directed evolution of gene families.

17. Wolfson AJ, Kanaoka M, Lau F, Ringe D, Yound P, Lee J, Blumenthal J: **Modularity of protein function: chimeric interleukin 1 betas containing specific protease inhibitor loops retain function of both molecules.** *Biochemistry* 1993, 32:5327-5331.

18. Campbell RK, Bergert ER, Wang Y, Morris JC, Moyle WR: **Chimeric proteins can exceed the sum of their parts: Implications for evolution and protein design.** *Nat Biotechnol* 1997, **15**:439-443.

19. • Crameri A, Dawes G, Rodriguez E, Silver S, Stemmer WPC: **Molecular evolution of an arsenate detoxification pathway by DNA shuffling.** *Nat Biotechnol* 1997, **15**:436-438.

This is the first paper to describe the use of DNA shuffling to evolve a multi-gene operon or gene cluster.

20. Moore JC, Arnold FH: **Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents.** *Nat Biotechnol* 1996, 14:458-467.

21. • Arnold FH, Moore J: **Optimizing industrial enzymes by directed evolution.** *Adv Biochem Eng Biotechnol* 1997, **58**: 2-14.

This paper describes the evolution of an industrial enzyme using both asexual random point mutation and DNA shuffling strategies.

22. Liu DR, Magliery TJ, Pastrnak M, Schultz PG: **Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*.** *Proc Natl Acad Sci USA* 1997, in press.

23. Mathias P, Wickham T, Moore M, Nemerow G: **Multiple adenovirus serotypes use αv integrins for infection.** *J Virol* 1994, **68**:6811-6814.

24. Bergelson JM, Cunningham JA, Droguett G, Kurt-Jones EA, Krithivas A, Hong JS, Horwitz MS, Crowell RL, Finberg RW: **Isolation of a common receptor for Coxsackie B viruses and adenoviruses a and 5.** *Science* 1997, **275**:1320-1323.

25. Stott J, Almond N: **Assessing animal models of AIDS.** *Nat Med* 1995, 1:295-297.

26. Wei-Shau H, Temin HM: **Retroviral recombination and reverse transcription.** *Science* 1990, **250**:1227-1233.

27. Ulmer JB, Donnelly JJ, Liu MA: **Toward the development of DNA vaccines.** *Curr Opin Biotechnol* 1996, **7**:653-658.

28. Stemmer WPC, Crameri A, Ha KD, Brennan TM, Heyneker H: **Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides.** *Gene* 1995, **164**:49-53.

29. Henco K, Brosius J, Fujisawa A, Haynes JR, Hochstadt J, Kovacic T, Pasek M, Schambock A, Schmid J, Todokoro K *et al.*: **Structural relationship of human interferon alpha genes and pseudogenes.** *J Mol Biol* 1985, **185**:227-260.

30. •• Zhang J, Dawes G, Stemmer WPC: **Evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening.** *Proc Natl Acad Sci USA* 1997, **94**:4504-4509.

This paper illustrates how visually screening a modest number of colonies (rather than using selection), combined with DNA shuffling, can give substantial improvements in a relatively large enzyme within a small number of cycles.

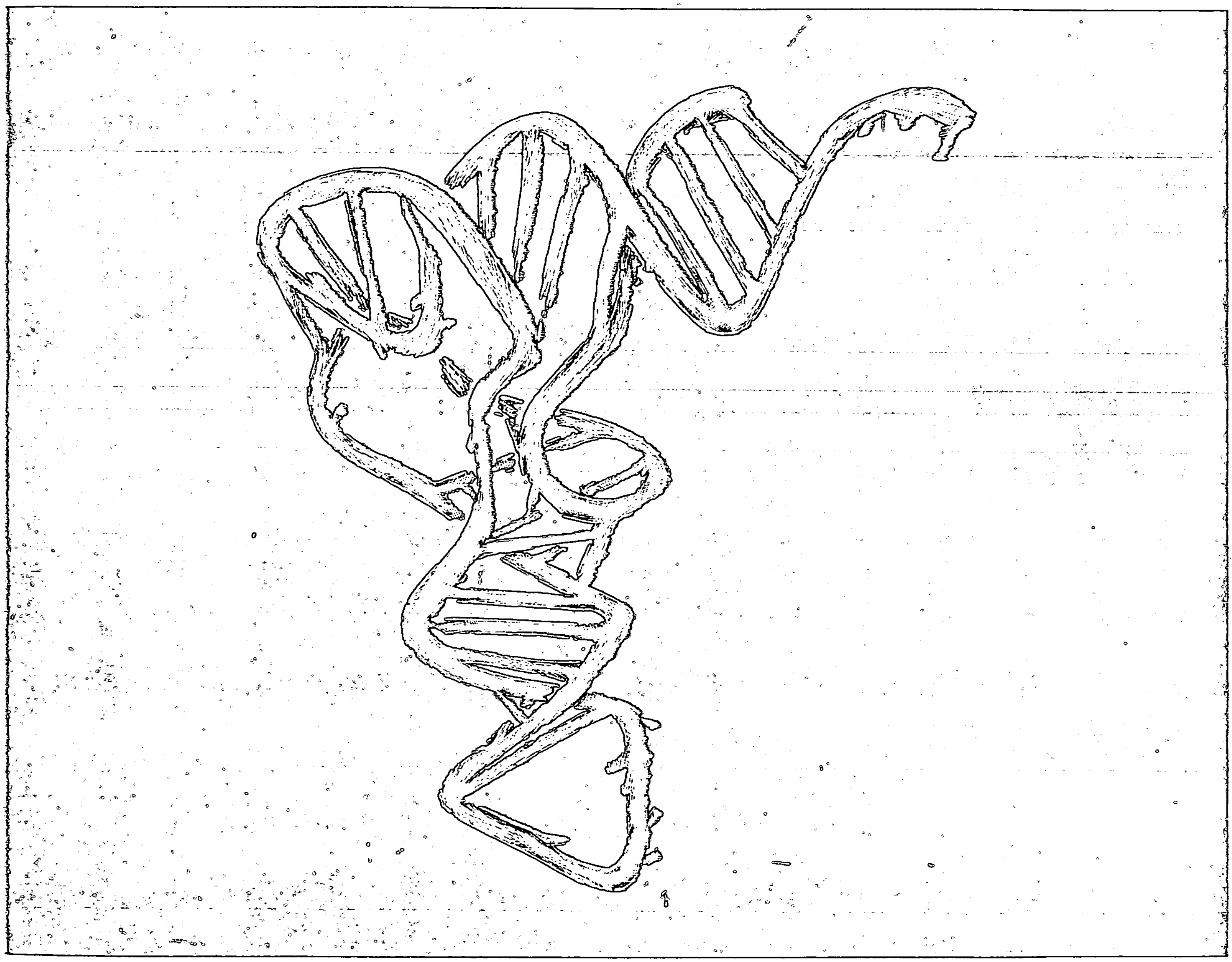
31. Crameri A, Whitehorn E, Tate E, Stemmer WPC: **Improved green fluorescent protein by molecular evolution using DNA shuffling.** *Nat Biotechnol* 1996, 14:315-319.

32. Crameri A, Cwirla S, Stemmer WPC: **Construction and evolution of antibody-phage libraries by DNA shuffling.** *Nat Med* 1996, 2:100-102.

2009 - 2209

VOLUME 20 Sequences Supplement MAY 11, 1992

Nucleic Acids Research



IRL PRESS
at
OXFORD UNIVERSITY PRESS
Oxford New York Tokyo

ISSN 0305 1048 Coden NARHAD

© 1992 Oxford University Press Nucleic Acids Research, Vol. 20, Supplement 2111–2118

Codon usage tabulated from the GenBank genetic sequence data

Ken-nosuke Wada[1], Yoshiko Wada[2], Fumie Ishibashi, Takashi Gojobori and Toshimichi Ikemura
National Institute of Genetics, Mishima, Shizuoka-ken 411, [1]Chukyo University, School of Computer and Cognitive Sciences and [2]Niigata University, School of Medicine, Japan

Codon usages in 22361 genes can be analyzed using the nucleotide sequence data obtained from the GenBank Genetic Sequence Data Bank (Release 69.0, Sept.,1991). The database is called as the CUTG Database (1–4), and is distributed on EMBL CD-ROM (December 1991; CODON by Wada et al.) as a member of NAR Sequence Supplement Databases (5). The CUTG codon database is also available for on-line access to DDBJ (DNA DATA BANK OF JAPAN): Please, contact with DDBJ (Mail Address; ddbj@ddbj.nig.ac.jp).

Files named as ***.CODON.69 list the codon use in each of gene registered in the GenBank Sequence files (gb***.seq). The LOCUS names given in the GenBank were used for designating individual genes, and the SHORT DIRECTORY of the GenBank Short Directory File is presented for defining each LOCUS name analyzed here (see ***.SDR.69).

To reveal the characteristics of the codon use of a wide range of organisms, as well as viruses and organella, the frequency (per one thousand) of codon use in each organism for which more than 20 genes are available was calculated by summing up numbers of codon use (***.total.69); Table 1 of this paper. The number of genes summed for each organisms is given in the row designated as GENES, and the total codon number thus summed is given at the bottom row. Names of the organisms are listed in the SPECIES FILE (Table 2). Amino acids are added simply according to the universal codon table.

METHODS

In selecting protein coding sequences we relied on the FEATURES tables of the GenBank, and only complete genes, starting with an initiation codon and ending with one of stop codons, were used in the analysis (see REFERENCES for details). In the GenBank, a group of consecutive genes whose entire region had been sequenced were registered under one LOCUS name. To distinguish the different genes belonging to a single LOCUS, symbol # followed by a number is added after the LOCUS name; the numbers represent the order of the peptides registered in the FEATURES of the GenBank. When introns of a gene have not been completely sequenced, some of its exons are registered in separate entries (LOCUS) in the GenBank. These exons belonging to the same gene but having different LOCUS names were combined, and the LOCUS name of the last exon followed by symbol * was given to the gene thus combined. The order of the codons in the table is the same as the previous compilation (see the CODON_LABEL file or REFERENCES).

ACKNOWLEDGMENTS

The authors are very grateful to Dr. S. Osawa (Nagoya Univ.). This work was supported by grant-in-aid for scientific research from the Ministry of Education, Science and Culture of Japan (No. 01656004).

REFERENCES

1. Maruyama, T., Gojobori, T., Aota, S. and Ikemura, T. (1986) Nucl. Acids Res. 14, r151–197.
2. Aota, S. Gojobori, T., Ishibashi, F., Maruyama, T. and Ikemura, T. (1988) Nucl. Acids Res. 16, r315–402.
3. Wada, K., Aota, S., Tsuchiya, R., Ishibashi, F., Gojobori, T., and Ikemura, T. (1990) Nucl. Acids Res. 18, Supplement, 2367–2441.
4. Wada, K., Wada, Y., Doi, H., Ishibashi, F., Gojobori, T., and Ikemura, T. (1991) Nucl. Acids Res. 19, Supplement, 1991–1986.
5. Wada, K., Wada, Y., Ishibashi, F., Gojobori, T., and Ikemura, T. EMBL CD-ROM (December 1991).

Table 1.

/		CHP	HUM	CRU	GPI	HAM	MUS	RAT	BOV	DOG	PIG	RAB	SHP	CHK	DUK	ONH	SMO	XEL	APL	BMO	CEL	CHI	DDI	DRO	MOT	PFA	SCM	SUP	SUS	TRB	ATH
No.GENES		23	2681	40	21	37	1294	1209	375	66	117	178	69	368	20	23	33	218	22	38	50	36	96	550	20	86	46	21	48	95	83
ARG	CGA	5.9	5.6	6.9	5.6	4.8	6.0	6.1	5.3	5.2	6.1	4.8	5.0	4.7	1.2	5.4	3.7	6.6	11.9	4.3	6.6	1.0	0.6	6.8	1.5	1.5	8.0	11.5	7.0	6.7	4.0
	CGC	13.0	11.2	8.5	12.7	13.7	10.5	10.8	11.4	8.9	11.7	14.1	9.9	12.4	24.4	10.5	19.5	9.2	13.8	9.2	10.5	1.0	0.1	18.1	6.0	0.2	2.6	29.7	8.0	14.8	3.7
	CGG	10.5	10.7	8.4	10.1	10.6	9.2	9.5	11.2	10.0	11.6	12.1	7.2	8.4	7.1	5.1	5.0	6.1	17.2	3.2	2.6	0.2	0.1	7.2	1.3	0.1	1.5	2.8	2.3	8.6	2.5
	CGU	3.1	4.6	4.7	4.8	7.9	4.6	5.2	3.8	3.3	4.1	3.9	3.5	6.3	7.8	2.7	13.3	9.2	8.9	8.8	20.6	18.1	17.4	9.5	10.0	3.2	16.6	33.7	17.6	14.3	12.4
	AGA	11.6	9.8	12.5	7.3	9.1	11.2	9.7	9.7	10.6	9.5	8.3	10.2	11.0	7.3	17.5	12.1	14.5	22.9	13.6	16.4	8.3	19.9	4.6	8.3	15.9	11.9	7.9	8.7	5.8	14.9
	AGG	17.1	10.8	7.8	8.5	10.0	11.2	10.6	11.6	11.3	10.5	10.7	11.6	10.7	8.8	15.6	16.5	12.3	21.0	8.4	1.8	0.2	1.0	5.8	6.2	2.7	4.3	10.7	9.4	7.7	13.4
LEU	CUA	6.9	6.1	8.4	6.7	4.8	7.2	6.8	5.0	4.9	5.6	4.5	5.1	4.9	3.7	10.5	4.8	7.7	7.5	4.9	2.9	2.9	1.9	6.4	7.3	3.3	7.6	5.5	6.3	8.5	7.9
	CUC	20.3	20.1	17.6	24.9	18.9	20.0	19.9	21.0	25.3	21.7	24.4	31.9	16.9	20.8	21.0	16.6	12.7	32.6	21.6	23.6	25.5	13.0	12.8	13.0	2.3	5.9	36.0	20.0	17.8	18.9
	CUG	52.8	42.1	32.5	50.3	38.5	40.3	42.1	45.8	45.0	46.6	50.9	49.7	40.4	56.7	47.4	51.9	25.8	28.9	19.2	6.6	0.5	0.6	37.8	19.5	0.4	6.4	13.9	15.3	19.9	7.2
	CUU	9.3	10.8	13.3	9.3	11.9	11.3	11.3	10.4	11.2	9.1	9.1	11.9	10.9	9.0	9.4	3.2	15.5	18.2	10.5	28.8	21.2	12.7	7.5	11.7	9.7	18.1	18.2	15.4	23.5	25.6
	UUA	2.0	5.4	5.9	2.9	3.1	4.8	4.9	4.4	4.9	4.8	4.3	4.0	4.8	2.9	3.5	2.0	8.1	6.5	9.2	2.6	6.0	44.8	4.0	7.8	49.0	27.9	0.0	3.7	7.0	6.2
	UUG	8.4	11.1	15.4	11.0	10.1	11.7	11.9	10.6	10.7	10.8	9.2	10.0	10.8	7.1	12.1	8.3	12.9	17.4	14.8	16.7	20.8	9.8	14.0	14.3	7.9	13.8	11.1	10.4	17.1	19.5
SER	UCA	5.9	9.7	9.9	6.4	8.2	10.5	9.1	7.9	7.1	7.9	6.7	8.3	8.6	5.4	10.8	6.3	11.6	17.6	5.6	11.0	51.7	32.5	6.6	7.5	24.4	18.9	7.5	10.4	10.9	12.0
	UCC	21.6	17.8	13.9	18.0	20.0	18.1	17.6	16.9	18.2	19.5	18.7	20.5	16.1	15.4	37.9	21.0	14.9	18.0	14.3	13.2	2.6	6.0	20.1	13.8	6.1	6.0	18.2	15.8	11.6	10.1
	UCG	4.1	4.1	2.6	3.3	4.2	4.3	4.0	4.5	4.5	4.2	4.4	4.7	5.5	6.6	2.4	2.3	3.9	16.2	9.2	9.2	0.3	1.6	17.0	5.5	1.4	4.3	1.2	4.4	8.2	5.4
	UCU	16.6	13.3	12.1	10.5	14.6	15.3	13.9	11.7	10.9	11.1	9.0	11.2	12.2	11.7	36.3	25.8	17.5	24.3	10.0	15.4	4.3	18.3	6.5	15.0	16.6	11.3	11.9	12.5	11.0	18.4
	AGC	18.9	18.7	14.7	20.6	14.7	19.9	17.2	17.6	16.1	17.8	19.4	21.7	19.5	26.4	18.0	17.8	16.7	18.0	17.6	5.8	7.7	2.4	19.5	12.0	4.4	7.0	20.2	13.5	13.4	11.5
	AGU	5.9	9.9	11.7	8.7	9.0	10.4	9.9	8.6	9.2	7.8	7.3	7.0	8.3	6.6	7.5	9.2	12.4	18.2	8.7	5.8	2.6	12.7	9.9	4.0	24.8	19.4	7.1	8.2	13.7	10.2
THR	ACA	10.5	14.3	15.5	12.2	15.9	15.3	14.6	12.2	13.2	11.8	10.5	10.3	14.7	4.2	11.0	5.5	18.3	15.2	10.7	10.8	24.5	21.0	10.3	6.2	24.5	22.0	9.5	13.7	17.7	15.3
	ACC	29.9	22.6	20.8	25.8	26.0	22.4	21.1	20.3	21.0	23.2	22.7	27.6	19.3	25.2	40.6	35.6	17.3	11.7	19.4	17.9	6.7	14.9	23.9	8.2	6.7	9.2	33.3	25.9	13.2	17.3
	ACG	6.6	6.6	4.2	9.2	4.8	6.3	5.9	7.7	7.2	7.9	8.6	7.0	7.6	8.8	6.2	2.0	4.0	9.3	6.1	6.2	0.5	0.7	13.9	4.5	2.4	7.3	4.8	6.0	15.9	6.2
	ACU	10.8	12.6	14.4	12.1	12.0	13.5	12.9	10.7	13.1	10.6	9.4	9.6	12.6	12.2	7.5	10.2	17.2	6.5	13.4	15.5	17.0	30.1	8.9	9.8	19.7	18.6	7.1	10.8	12.4	19.9
PRO	CCA	11.6	15.4	10.8	10.5	14.5	15.6	14.3	12.9	11.4	11.9	12.0	10.3	13.1	6.6	10.8	8.5	18.1	8.1	7.5	37.8	22.6	40.1	12.9	14.0	26.8	14.2	10.3	20.7	11.8	17.3
	CCC	20.2	20.6	11.0	22.2	19.0	18.9	17.6	19.9	18.5	21.5	22.5	24.8	18.3	20.8	15.1	18.8	13.6	10.3	14.4	3.0	2.1	0.5	19.1	15.2	3.9	4.4	17.0	13.5	11.0	7.1
	CCG	9.3	6.8	2.3	6.0	3.4	6.2	5.6	7.6	5.3	7.2	8.9	6.3	7.4	6.8	17.0	7.8	4.0	10.1	7.4	6.6	0.5	0.4	16.4	4.8	0.6	3.2	1.6	5.1	9.3	6.9
	CCU	11.2	16.1	12.1	11.2	18.3	17.2	15.7	14.4	12.6	14.8	13.2	14.4	12.5	8.3	11.3	13.0	15.3	12.5	11.1	5.5	2.8	3.4	6.2	13.5	11.6	8.3	9.9	13.0	11.4	16.8
ALA	GCA	11.8	14.4	20.1	12.5	12.6	14.0	14.9	13.5	13.8	12.6	11.5	6.2	17.5	9.3	15.1	11.5	19.8	15.8	9.2	12.5	26.9	16.6	12.5	11.8	24.7	17.2	22.6	19.0	23.4	14.9
	GCC	30.7	29.7	26.9	30.7	27.1	26.6	28.5	34.0	29.8	31.8	33.9	33.3	26.6	49.3	28.5	36.3	20.0	15.2	26.5	25.6	40.6	13.6	37.9	25.5	6.5	7.0	52.3	34.4	18.9	16.1
	GCG	18.5	7.2	4.9	7.3	5.6	6.6	6.6	8.7	6.6	8.5	8.4	8.4	9.2	16.6	16.4	12.5	4.2	7.5	11.0	4.8	0.2	0.3	14.3	14.3	0.5	3.5	2.4	6.2	18.7	7.6
	GCU	19.8	18.9	27.1	18.0	21.7	19.3	20.9	18.1	18.8	18.7	15.3	17.0	23.9	30.0	22.9	20.5	23.5	18.0	24.2	32.4	85.2	25.1	16.2	29.0	23.5	20.2	33.7	25.6	21.2	37.5
GLY	GGA	15.7	17.4	24.7	12.4	17.5	17.1	16.2	15.8	17.5	14.0	14.8	16.3	16.4	13.7	13.5	14.3	22.7	12.5	26.6	56.2	41.3	6.7	19.4	12.8	26.3	25.1	35.6	35.5	16.0	30.0
	GGC	24.6	25.3	22.4	27.2	25.8	23.5	23.4	26.6	24.0	30.5	27.8	25.9	22.4	30.3	25.8	32.5	16.1	19.2	31.7	4.0	2.8	2.6	31.3	18.2	1.5	18.4	30.5	21.4	17.8	10.1
	GGG	25.6	17.5	15.4	17.1	12.7	15.9	15.6	17.5	16.0	17.4	17.8	14.9	15.1	12.0	10.2	12.3	12.7	10.9	6.0	2.3	0.3	0.9	4.6	3.3	2.1	4.5	6.3	8.0	11.3	7.4
	GGU	6.6	11.5	14.7	7.5	15.2	11.4	11.6	11.0	10.9	11.5	8.3	10.6	11.3	8.3	5.1	9.3	15.3	9.9	38.2	11.6	17.2	56.8	16.0	14.0	28.4	55.6	25.3	26.8	27.5	29.4
VAL	GUA	4.6	6.1	7.1	5.7	5.8	6.6	6.6	5.6	6.0	5.2	4.5	3.9	6.4	3.7	5.9	3.8	9.4	13.4	10.7	4.5	3.4	12.8	5.3	13.2	20.2	14.8	4.8	8.5	11.8	4.7
	GUC	13.8	16.2	20.5	18.9	15.5	16.9	16.8	17.2	18.8	18.6	18.4	15.5	15.5	14.7	21.8	23.1	12.4	15.4	23.3	19.1	30.0	11.7	15.4	26.9	3.1	11.9	35.6	25.0	11.3	19.3
	GUG	41.2	30.7	29.1	39.2	34.4	30.4	31.4	33.0	32.1	33.2	32.9	31.4	29.6	32.0	23.7	27.1	21.3	12.3	17.4	8.8	0.9	1.7	26.0	19.9	2.6	12.6	14.3	11.4	27.0	18.1
	GUU	8.5	10.2	15.8	7.4	11.0	9.6	9.9	9.9	11.5	9.2	8.6	6.3	12.2	11.5	4.6	7.2	15.9	9.5	14.8	22.9	43.1	36.0	10.4	22.7	22.7	21.8	9.9	16.4	22.6	28.4
LYS	AAA	11.8	21.9	29.9	19.5	17.0	21.5	21.2	22.4	21.9	20.7	18.4	18.2	26.2	21.0	11.0	15.8	34.1	37.0	21.7	20.1	36.2	57.5	15.0	32.0	81.3	48.0	23.0	18.7	24.4	22.3
	AAG	38.9	35.2	35.4	34.6	39.5	38.2	39.2	37.3	33.6	37.4	36.4	38.4	41.9	78.7	29.6	59.6	35.6	21.5	30.8	48.5	42.4	15.9	41.1	44.5	15.1	17.6	93.1	52.8	35.5	40.7
ASN	AAC	16.2	22.3	21.3	26.9	25.2	23.1	23.4	22.9	24.7	24.0	25.4	24.8	24.2	21.5	27.7	31.0	21.2	31.8	32.4	26.3	24.3	15.4	29.1	43.7	20.7	12.5	30.1	25.1	23.5	25.4
	AAU	6.6	16.5	20.8	13.6	13.2	15.9	15.1	13.4	16.7	14.5	12.3	9.8	15.2	8.1	4.6	6.3	20.1	12.8	14.3	20.8	15.3	42.5	20.8	12.3	77.6	39.9	1.6	11.1	18.8	14.5
GLN	CAA	7.9	10.8	8.4	11.2	8.9	11.2	10.5	10.2	11.2	8.1	8.3	10.5	10.8	5.9	10.0	7.0	15.7	19.6	12.8	36.3	30.0	43.0	15.1	16.9	28.4	23.4	8.3	16.0	13.6	17.4
	CAG	57.2	33.8	31.6	33.1	35.3	32.8	32.1	34.1	32.1	34.4	32.2	39.3	32.2	29.6	19.1	27.6	26.9	16.8	17.5	16.0	2.8	1.0	39.7	19.0	2.6	6.0	26.5	22.8	20.0	19.9
HIS	CAC	18.7	14.7	9.1	18.3	13.3	14.8	14.0	15.0	13.5	14.1	17.5	15.4	14.0	18.6	12.1	13.6	12.9	12.8	11.1	11.1	13.1	5.7	16.2	17.4	7.1	5.7	10.3	11.4	13.1	11.2
	CAU	10.3	9.3	7.7	7.6	9.5	9.4	8.5	7.5	7.7	7.1	7.1	7.6	8.0	7.3	5.4	6.2	11.9	3.8	4.5	11.8	12.1	12.0	10.3	5.0	18.7	19.2	7.5	10.2	9.2	8.2
GLU	GAA	13.4	26.4	31.3	16.5	22.6	26.0	26.7	26.6	26.0	24.0	22.7	18.2	31.2	21.3	24.5	16.8	37.2	32.8	27.1	40.8	31.3	57.8	18.1	30.0	77.1	46.1	13.5	25.2	33.8	29.1
	GAG	60.5	41.6	31.2	42.9	44.2	41.2	43.2	45.4	36.8	40.4	48.0	39.0	48.7	39.1	28.3	39.6	35.5	27.5	26.2	41.8	6.2	8.0	44.3	32.4	7.7	12.5	30.9	32.9	36.9	35.0
ASP	GAC	33.0	28.9	26.2	33.3	32.4	28.3	29.6	31.2	29.0	29.3	34.7	30.8	27.8	26.1	61.6	45.9	25.2	40.5	32.6	23.7	27.0	7.9	24.6	45.9	8.0	14.0	13.1	28.7	26.5	24.4
	GAU	14.8	21.5	23.1	18.9	20.9	21.0	21.5	20.4	21.3	20.4	18.5	14.8	26.1	15.4	12.6	17.6	29.7	27.7	21.3	36.4	26.3	51.0	26.4	18.7	52.8	46.6	5.5	27.9	27.5	32.2
TYR	UAC	18.4	16.0	15.3	23.6	20.5	18.2	17.9	20.3	19.9	19.6	20.7	20.6	18.2	21.0	17.0	17.8	16.1	15.2	38.6	15.4	11.0	9.9	18.4	52.6	5.6	14.8	26.1	21.9	15.2	19.6
	UAU	9.2	12.3	13.6	9.0	10.3	12.3	11.8	10.8	11.7	9.6	9.5	9.4	10.0	5.1	5.9	4.7	14.4	9.5	11.3	11.3	6.2	23.5	9.9	12.8	31.9	36.8	2.4	7.0	10.8	9.2
CYS	UGC	10.5	13.8	11.7	16.7	12.1	12.7	12.0	13.5	16.6	17.2	11.6	29.2	12.8	6.6	21.8	23.0	12.9	8.1	20.8	9.9	5.3	2.5	14.6	9.0	2.2	4.8	1.2	13.8	10.6	8.1
	UGU	5.9	9.9	7.9	7.5	7.0	10.6	9.7	9.1	12.0	10.5	6.4	12.9	7.7	3.4	15.1	12.5	11.3	6.5	14.2	9.5	2.9	13.7	5.4	4.0	13.8	15.0	0.4	8.3	8.0	7.6
PHE	UUC	21.0	22.1	21.2	24.9	28.1	23.2	26.1	26.0	26.5	24.2	32.1	29.0	20.4	23.9	18.0	23.1	16.5	27.5	29.0	22.2	56.1	18.1	22.2	57.2	9.6	17.5	15.4	22.1	18.6	27.0
	UUU	7.2	15.8	24.1	14.6	15.9	15.2	17.0	15.6	17.6	14.8	16.4	13.9	14.3	12.9	6.7	7.2	18.7	16.6	10.5	9.3	12.6	22.0	11.3	9.7	25.2	22.3	2.8	10.7	18.6	15.3
ILE	AUA	3.9	6.1	10.2	6.2	4.8	6.1	6.1	5.5	7.4	5.5	5.2	5.0	6.7	4.9	3.8	2.7	9.5	6.5	8.3	2.7	2.1	6.9	7.4	5.2	25.2	23.9	0.4	5.3	9.9	8.2
	AUC	19.7	24.3	30.0	32.9	27.1	24.5	27.7	26.3	30.3	28.0	30.9	25.5	25.9	32.2	24.8	28.3	18.8	20.2	28.1	25.1	33.4	20.5	25.2	25.2	4.6	12.4	42.4	36.5	16.3	27.6
	AUU	8.9	15.0	26.4	14.9	17.1	14.6	16.3	13.2	15.0	16.1	13.6	11.2	16.4	12.5	8.3	7.3	20.1	7.1	16.8	21.3	24.3	44.0	15.6	14.0	28.8	27.1	5.9	11.4	20.3	24.0
MET	AUG	16.9	22.3	25.4	21.8	27.3	22.8	25.3	23.8	23.9	21.9	24.9	28.6	24.3	21.5	32.6	25.5	24.8	29.1	31.0	22.8	17.7	18.5	23.4	26.4	18.2	32.2	17.4	32.1	27.2	27.0
TRP	UGG	20.2	13.7	9.3	14.4	11.7	13.7	12.4	12.7	15.3	15.0	13.5	10.1	10.9	5.6	7.8	4.5	9.9	6.1	12.4	7.7	11.7	9.0	9.4	9.8	3.9	7.1	0.0	9.5	8.6	11.4
TER	UAA	1.1	0.7	0.7	0.7	0.9	0.7	0.7	0.9	1.2	0.7	0.5	0.6	1.0	1.5	1.3	1.7	1.4	3.2	1.9	1.0	5.9	2.4	1.0	2.0	1.4	2.0	5.9	2.0	1.1	1.1
	UAG	0.5	0.5	0.4	0.6	0.5	0.5	0.6	0.6	0.6	0.5	0.4	1.3	0.5	1.0	3.5	2.2	0.5	0.0	1.0	0.3	0.0	0.1	0.5	0.3	0.1	0.4	2.0	1.1	0.4	0.5
	UGA	2.1	1.1	1.2	1.2	1.4	1.4	1.2	1.3	2.0	1.5	0.9	2.5	0.8	2.4	1.3	1.7	0.8	1.2	1.0	0.2	0.3	0.2	0.4	1.0	0.2	1.2	0.4	0.3	0.4	1.2
	TOTAL	6097	1145022	17238	8378	13044	492365	490242	131073	17799	42948	100704	15635	163534	4094	3716	6009	83168	4941	9438	32412	5807	35984	299748	5994	47407	12930	2525	14177	49564	29154

2		BLY	BNA	CRE	EME	MZE	NEU	PEA	PHV	POT	RIC	SLM	SOY	SPI	TOB	TOM	WHT	YSA	YSC	YSK	YSP	ACC	AFA	ANA	ATU	AVI	BAC	BPE	CHT	CLO	COR
No.GENES		72	22	37	32	129	69	47	28	34	55	21	65	39	56	59	64	27	891	30	81	30	40	47	70	51	636	26	62	74	21
ARG	CGA	1.3	4.6	0.2	9.8	3.4	3.0	4.3	3.0	5.2	2.6	2.0	4.5	4.0	4.8	4.4	2.6	0.3	2.3	1.0	7.3	5.4	4.0	4.0	11.0	0.3	5.1	2.0	5.2	0.9	4.4
	CGC	15.2	6.7	42.3	14.9	17.1	21.6	3.5	5.0	2.3	14.2	14.1	7.5	3.8	3.8	3.1	8.0	0.3	2.1	0.8	6.8	7.7	34.4	14.2	24.3	38.0	8.1	35.0	6.6	1.8	36.1
	CGG	11.6	2.2	3.3	9.7	8.5	5.1	1.2	2.1	1.5	7.6	1.3	2.2	3.0	2.2	1.9	5.9	0.1	1.3	1.0	2.9	2.3	12.6	9.2	14.8	6.4	4.4	13.4	2.0	1.4	3.2
	CGU	8.2	10.1	5.0	11.2	6.2	12.3	8.5	8.4	8.9	7.2	13.9	7.4	7.8	7.7	9.3	2.4	3.3	7.2	2.8	19.5	24.8	9.5	17.6	14.3	11.2	10.1	5.2	12.7	3.6	17.1
	AGA	2.4	11.7	0.0	5.5	5.4	4.5	16.6	10.9	8.7	8.1	6.2	14.8	13.6	12.5	14.8	4.5	40.2	23.2	24.8	11.2	2.1	1.5	7.8	6.5	0.4	12.6	1.0	15.1	20.8	0.6
	AGG	11.7	12.4	0.3	6.4	15.4	7.7	10.3	11.7	10.2	12.5	4.2	11.3	15.1	12.1	10.0	8.7	0.8	8.1	3.6	4.5	0.1	2.3	2.1	7.9	1.6	3.6	3.9	2.9	5.1	2.2
LEU	CUA	3.1	8.9	0.5	10.1	8.4	3.2	6.4	8.9	8.0	6.1	3.4	7.0	8.3	7.4	7.2	10.0	0.5	12.0	10.8	7.1	6.6	2.7	14.3	8.6	0.8	7.1	0.6	9.8	7.1	4.2
	CUC	31.9	20.9	6.6	22.1	25.7	36.3	14.1	17.9	10.6	30.0	34.8	16.2	13.2	12.7	11.5	21.0	3.0	4.3	2.6	7.4	9.9	17.3	11.2	21.9	19.9	7.6	13.2	7.5	3.1	24.8
	CUG	23.1	9.7	69.1	21.0	27.3	16.4	6.9	12.3	7.5	17.8	7.6	10.3	8.6	7.1	6.1	16.1	1.1	9.0	3.8	6.1	11.8	47.9	15.5	27.4	67.0	14.1	57.0	5.5	6.3	30.8
	CUU	7.7	21.4	2.2	18.3	16.2	15.0	23.6	22.6	26.9	11.5	13.8	23.7	21.8	21.8	25.0	9.2	5.3	10.1	10.2	26.9	16.1	9.1	6.1	19.5	1.8	19.2	4.3	17.0	17.1	16.3
	UUA	0.9	9.3	0.0	7.1	4.8	0.8	9.2	8.8	14.1	5.0	3.9	8.4	8.7	10.3	10.3	2.1	18.9	24.5	38.9	23.8	27.8	1.9	23.9	4.6	0.2	25.8	0.4	25.5	32.5	3.2
	UUG	5.0	12.8	0.6	13.9	13.2	12.5	22.4	16.5	22.7	10.3	17.8	20.5	23.6	20.7	24.6	13.6	57.1	31.0	17.4	24.0	15.5	12.3	24.9	16.7	4.3	12.9	10.7	19.8	10.3	14.5
SER	UCA	3.4	8.9	0.4	12.3	8.9	4.1	16.0	12.8	16.6	7.0	5.7	14.7	16.9	15.8	16.7	13.9	7.4	16.2	15.9	17.1	13.5	4.0	8.7	9.5	0.1	13.7	1.0	8.4	19.0	5.0
	UCC	20.0	13.6	21.4	19.5	16.5	22.2	14.7	20.6	8.1	13.8	28.5	9.5	15.5	10.0	8.7	12.9	22.1	14.9	6.9	12.2	3.0	10.1	9.9	13.9	18.0	8.3	14.8	12.0	4.6	24.8
	UCG	9.9	6.0	21.8	12.1	10.4	12.2	4.0	4.1	3.2	9.4	6.7	4.3	5.0	4.3	3.8	8.8	2.3	7.1	4.0	7.0	5.0	19.1	3.0	16.5	10.7	6.5	20.3	5.0	2.7	5.8
	UCU	4.2	14.4	3.8	16.6	10.0	11.5	21.7	23.8	15.9	8.8	16.5	17.3	19.0	20.6	17.4	16.2	21.0	24.6	29.6	32.3	12.2	2.2	16.0	8.6	0.9	14.9	1.2	29.2	15.1	12.8
	AGC	16.5	12.0	13.1	14.9	17.1	14.5	10.6	15.7	8.4	17.4	9.7	16.8	11.3	9.0	10.2	13.8	1.9	7.9	3.9	8.9	8.0	15.0	10.8	17.1	18.9	11.5	27.9	10.4	8.3	10.5
	AGU	3.6	8.4	0.4	8.0	4.4	4.7	13.9	10.8	13.4	8.8	3.7	14.0	10.5	11.9	12.1	3.9	4.7	12.1	17.6	13.1	10.0	3.7	9.1	8.2	2.9	10.1	2.4	8.2	15.8	2.2
THR	ACA	4.8	16.2	0.8	14.2	8.5	6.7	16.9	12.7	19.4	8.1	12.6	14.9	18.4	16.4	18.2	6.8	5.6	16.1	17.5	13.4	12.1	4.8	18.3	9.7	0.3	24.1	1.6	25.5	24.7	4.5
	ACC	27.8	17.3	45.9	20.1	20.0	31.4	13.4	18.9	9.9	20.5	27.2	14.0	19.8	11.9	11.7	17.5	28.7	13.7	8.0	11.7	15.4	26.8	21.4	18.8	39.4	8.4	29.2	9.1	8.6	35.9
	ACG	14.0	9.2	6.9	10.8	11.4	9.1	4.1	6.7	3.2	8.1	5.4	3.2	3.8	4.0	3.8	6.1	0.8	6.9	3.1	5.3	10.1	18.8	2.9	17.8	4.0	15.0	27.7	6.6	3.7	7.1
	ACU	4.9	16.5	5.9	16.5	8.4	11.3	24.0	13.2	29.2	10.3	17.1	17.4	16.0	20.4	20.1	12.7	29.8	21.9	22.9	24.0	18.9	5.0	14.6	7.3	2.6	13.1	3.8	26.8	23.2	10.5
PRO	CCA	12.0	17.7	0.4	13.4	14.9	6.4	23.3	13.5	18.8	10.0	9.1	30.6	18.9	26.2	20.9	56.0	34.9	20.8	15.5	11.6	21.3	6.2	11.1	10.9	0.9	12.1	1.9	11.0	14.5	18.4
	CCC	17.9	9.3	35.5	15.0	16.9	28.1	5.6	11.2	7.7	11.2	21.5	10.5	11.1	9.0	7.4	12.8	0.8	5.9	2.1	8.7	4.2	13.3	11.6	11.6	15.1	2.6	13.5	3.9	3.2	8.1
	CCG	17.5	7.8	5.7	11.6	18.5	11.0	5.7	2.8	3.0	14.8	1.7	4.6	5.7	3.5	2.6	17.9	1.2	4.5	2.4	4.4	10.0	24.0	2.5	19.3	30.0	12.6	22.1	2.6	4.8	8.0
	CCU	7.7	15.2	2.7	15.0	13.3	13.8	20.1	16.3	17.4	11.9	11.8	22.2	21.6	21.7	17.3	10.5	7.6	12.9	15.6	23.4	13.6	4.4	15.1	9.1	1.4	11.4	3.0	21.3	13.3	8.9
ALA	GCA	9.9	20.8	2.1	16.3	15.6	6.5	21.7	20.0	26.2	15.6	11.6	20.5	24.9	22.3	23.4	12.5	6.6	15.0	16.1	13.6	30.5	14.8	24.2	20.0	3.7	22.1	7.8	28.5	25.4	32.1
	GCC	46.1	20.0	64.7	27.1	34.5	48.7	11.5	22.0	10.7	30.3	32.9	16.6	18.8	16.2	13.9	24.2	21.9	14.7	6.0	13.2	14.5	46.6	16.9	32.2	65.6	12.3	65.4	10.3	9.4	30.2
	GCG	28.1	13.2	16.7	14.5	24.7	13.1	5.7	4.4	4.4	20.6	9.2	4.7	7.0	4.4	3.7	15.3	0.6	5.5	3.3	4.5	15.2	42.5	11.2	31.1	20.8	16.3	56.8	8.2	4.8	21.3
	GCU	12.9	33.9	21.5	24.1	27.1	25.3	38.5	29.4	38.6	17.3	30.6	23.2	35.5	35.2	38.0	13.4	44.1	26.5	21.9	33.2	21.5	10.9	38.5	19.5	6.6	20.2	7.1	48.1	25.2	27.9
GLY	GGA	11.7	28.0	0.5	15.2	11.0	7.5	29.7	41.8	31.7	17.1	39.6	22.7	23.8	31.4	32.0	28.3	6.2	9.4	18.3	16.0	11.0	7.3	10.4	12.1	2.5	22.6	7.1	30.0	33.6	13.6
	GGC	53.5	10.3	63.5	24.1	32.1	35.5	10.4	16.4	13.8	40.0	15.3	11.9	9.7	14.0	11.1	27.1	5.4	9.3	3.8	8.8	19.5	52.2	16.5	30.0	60.3	19.0	80.2	10.2	8.9	36.3
	GGG	20.2	10.1	0.9	11.1	14.0	5.3	7.8	15.4	6.8	14.7	5.5	11.0	14.3	10.0	9.2	31.0	2.4	5.3	2.6	4.1	5.6	11.5	5.5	12.5	5.0	11.4	17.5	8.1	5.1	4.6
	GGU	13.5	23.9	13.1	19.3	15.9	28.1	26.5	42.7	25.4	19.1	24.8	22.1	24.0	30.0	31.2	12.0	55.6	31.9	28.2	27.4	41.3	13.7	37.2	15.9	13.7	17.1	10.6	17.2	25.1	18.1
VAL	GUA	3.3	5.7	0.5	6.7	5.1	2.7	10.1	5.9	10.2	8.7	6.4	7.5	12.1	11.5	9.9	4.6	2.4	10.3	17.0	12.1	15.9	5.4	21.3	7.1	2.8	19.5	4.9	21.0	25.9	7.8
	GUC	28.0	20.8	26.1	24.3	22.1	37.8	10.1	11.7	9.5	27.8	23.2	9.0	14.1	13.8	10.6	16.9	23.3	13.9	6.9	13.5	10.6	30.0	10.7	22.5	32.6	13.8	35.8	9.0	5.2	27.0
	GUG	33.1	19.5	42.2	13.9	28.3	11.1	18.7	21.3	16.0	20.2	23.7	25.8	20.0	14.4	17.8	17.9	3.8	9.9	5.4	7.4	17.7	37.7	11.1	18.8	33.6	14.9	38.0	9.9	6.7	27.2
	GUU	7.5	28.7	4.7	17.1	12.5	16.8	35.0	23.5	32.8	15.5	23.8	24.4	32.8	29.4	31.7	12.1	35.2	25.0	23.6	29.4	24.1	8.4	24.2	17.0	4.8	20.3	6.2	30.9	25.6	24.0
LYS	AAA	6.7	28.0	0.2	15.3	9.7	5.0	33.2	26.9	31.7	8.8	10.7	23.2	28.0	23.2	30.3	5.9	28.9	39.2	64.4	35.6	39.9	6.7	37.6	14.2	6.5	48.4	4.4	56.2	54.9	8.5
	AAG	49.1	35.2	56.0	31.8	37.2	48.3	39.5	24.3	34.7	41.2	33.4	35.2	44.0	30.7	34.4	25.1	40.6	34.5	21.6	26.5	13.0	30.9	16.7	22.4	43.5	17.2	35.3	17.3	18.1	22.8
ASN	AAC	32.0	25.5	39.1	27.5	25.8	32.7	22.1	33.7	15.1	30.5	29.6	29.2	21.8	26.0	20.8	15.8	28.3	25.8	20.6	20.0	19.6	21.7	21.9	17.0	27.6	21.7	26.5	18.1	18.2	27.1
	AAU	5.4	10.6	0.9	13.8	9.2	7.5	22.8	19.6	31.7	14.3	10.1	20.2	18.1	27.8	27.8	7.1	15.6	32.9	50.4	32.7	28.8	9.9	17.3	15.2	4.8	32.1	10.4	26.4	50.5	7.6
GLN	CAA	16.0	35.7	1.5	16.1	22.7	9.1	19.7	20.6	20.0	23.1	20.7	27.4	19.9	22.5	21.4	130.6	35.5	29.2	21.8	27.2	35.0	7.4	37.1	14.7	4.4	25.9	8.1	26.5	18.7	10.5
	CAG	31.4	29.6	37.4	25.9	30.7	27.9	10.6	15.3	10.1	26.8	13.8	20.7	11.6	14.1	11.1	67.4	3.5	11.4	6.5	9.7	19.1	27.7	13.8	21.7	28.3	13.6	35.6	9.5	8.7	22.8
HIS	CAC	17.9	9.8	17.0	13.9	14.7	17.8	7.7	10.0	6.3	14.9	17.6	8.9	7.3	8.9	6.0	8.6	14.6	8.3	5.3	6.6	7.6	12.8	9.2	9.9	18.8	6.2	9.9	4.2	4.4	18.9
	CAU	4.8	6.7	1.2	11.9	7.2	6.6	12.5	11.4	14.7	7.9	8.9	12.1	8.9	11.1	9.0	6.8	8.0	12.8	13.4	15.6	17.6	8.5	6.0	11.4	5.6	14.6	7.4	7.7	11.2	4.0
GLU	GAA	7.4	27.7	0.6	23.7	15.0	10.3	37.6	28.3	38.2	13.9	19.0	34.6	28.4	26.7	32.2	10.5	58.0	48.4	52.8	42.8	41.2	25.5	58.2	30.3	33.9	49.5	16.9	43.9	46.3	40.0
	GAG	43.4	36.3	67.7	35.8	43.8	48.2	30.1	30.7	25.8	37.9	44.5	35.8	31.0	26.2	32.1	26.8	6.2	17.8	11.2	21.8	13.2	30.7	15.1	30.7	43.3	20.9	25.8	18.0	12.7	34.3
ASP	GAC	45.9	24.2	45.3	28.5	29.5	34.4	17.7	22.4	14.5	33.4	27.5	18.8	18.3	17.2	17.5	17.3	31.4	21.8	12.3	17.0	15.8	34.5	22.4	26.6	46.6	17.9	39.1	16.3	15.2	37.4
	GAU	9.7	23.1	9.3	24.7	14.4	19.6	31.5	24.6	37.8	17.4	22.5	29.7	30.6	33.0	33.5	7.8	30.6	37.7	49.9	38.9	41.6	21.7	31.1	29.3	13.8	37.8	18.8	37.1	48.4	32.5
TYR	UAC	25.6	14.7	29.0	18.3	21.6	24.5	12.5	18.5	16.7	24.1	26.2	17.1	15.3	16.6	16.4	24.4	22.6	16.0	12.7	12.2	10.5	16.5	18.6	8.8	21.2	12.4	14.9	11.9	10.8	19.6
	UAU	3.2	7.9	1.1	11.0	6.1	6.3	14.4	17.9	18.5	11.5	6.7	16.2	8.9	20.6	12.6	9.3	15.4	17.1	41.0	21.3	25.7	10.3	13.5	13.6	5.5	26.5	9.4	12.8	38.8	5.9
CYS	UGC	14.6	9.7	15.3	8.0	12.7	9.5	4.4	5.6	10.1	13.9	12.8	9.9	5.9	8.1	8.7	13.6	0.6	3.7	3.5	5.2	4.4	9.0	5.6	6.7	16.0	3.3	3.9	8.5	3.1	4.1
	UGU	4.2	9.7	0.5	3.9	4.3	2.2	8.5	3.9	11.9	5.3	4.7	5.5	9.0	10.1	8.5	4.1	9.5	7.8	13.8	9.0	7.9	2.2	6.0	3.0	2.1	4.0	0.3	10.2	7.1	1.2
PHE	UUC	32.0	24.1	37.9	24.6	29.5	29.9	20.4	25.0	15.2	30.8	30.2	23.5	19.0	18.1	19.5	15.3	25.8	19.3	16.9	13.2	13.1	24.7	14.9	21.2	31.2	13.4	17.9	19.0	10.4	23.4
	UUU	7.1	15.2	2.1	10.4	10.1	8.2	21.0	16.3	27.0	11.8	10.7	19.3	19.2	25.2	23.7	13.4	25.1	23.4	30.4	27.3	28.5	8.7	19.5	16.3	1.6	27.3	1.6	20.5	31.6	7.6
ILE	AUA	3.8	8.2	0.1	6.8	5.9	2.1	11.2	8.2	9.7	6.9	3.0	12.2	7.7	10.7	10.6	5.5	2.3	14.2	42.1	10.3	2.7	1.0	6.8	7.2	0.5	10.8	1.7	9.8	39.6	0.8
	AUC	36.0	20.6	36.8	27.0	25.7	32.8	17.3	21.1	14.4	28.4	37.3	15.0	19.1	13.7	17.4	19.2	21.6	18.0	10.8	14.0	16.1	41.2	25.4	28.7	50.0	21.7	29.4	22.7	9.3	33.6
	AUU	6.3	21.9	7.9	20.3	12.2	15.1	24.3	19.5	31.3	14.6	18.5	23.7	27.2	29.4	31.7	9.1	33.8	30.6	26.5	36.2	41.8	13.2	35.4	18.1	4.3	35.1	5.2	29.5	29.5	18.4
MET	AUG	28.1	27.1	27.9	23.0	24.0	23.3	19.5	14.5	25.0	23.5	31.1	20.1	23.3	22.8	26.4	16.1	20.5	21.3	21.8	23.0	23.1	25.0	19.7	25.4	28.5	22.2	14.2	21.6	24.3	17.5
TRP	UGG	14.4	8.9	10.4	12.4	11.3	14.3	11.1	12.7	9.6	15.9	10.4	11.5	12.1	13.7	13.2	8.3	12.2	9.9	12.7	8.1	15.6	13.2	10.2	10.0	10.1	12.2	7.5	10.5	13.6	9.2
TER	UAA	1.3	1.4	2.8	0.5	0.6	1.5	2.2	0.7	1.7	1.3	3.2	1.3	2.3	1.1	2.1	0.6	1.6	1.0	1.4	1.3	2.0	0.6	2.6	0.8	0.7	1.7	0.2	2.0	1.7	1.7
	UAG	1.0	1.4	0.7	0.6	1.2	0.4	0.5	1.0	0.5	0.8	0.0	1.0	0.7	0.4	0.5	0.4	0.5	0.4	0.1	0.3	0.3	0.4	0.9	0.9	0.1	0.4	0.4	0.5	0.4	0.6
	UGA	1.4	0.6	0.0	0.3	1.3	0.6	1.0	1.4	0.9	1.1	0.3	0.6	0.8	1.2	0.7	1.8	0.2	0.5	0.4	0.2	0.8	1.7	0.5	1.5	2.1	0.6	0.6	0.4	0.1	0.4
	TOTAL	19641	6311	10747	21864	42071	28412	12648	9009	11030	16639	5956	21723	10620	20020	17959	22540	11830	459247	16426	43812	9679	15113	11590	21321	18214	243715	20060	21215	32755	7791

		DVU	ECO	ERW	FDI	FPL	HAL	HEI	INS	KPN	LAC	MBI	MSG	MVA	NGO	PRM	PSE	R10	RCA	RHB	RHM	RSP	RSS	SHF	SMA	SSP	STA	STM	STR	STY	SYC
No.GENES		23	1562	54	31	39	76	33	26	123	51	20	23	22	31	31	259	23	51	41	126	25	24	20	36	33	73	140	102	223	23
ARG	CGA	3.3	3.1	2.5	4.2	3.7	8.0	5.9	11.5	2.4	6.8	0.9	4.4	1.4	1.4	5.4	4.4	6.8	1.3	5.5	5.7	3.5	1.0	2.8	1.7	3.0	3.3	3.4	3.2	3.5	2.8
	CGC	23.3	22.1	18.4	20.7	13.6	28.4	7.8	21.7	38.4	2.8	4.2	18.5	0.3	23.2	10.5	32.1	21.1	35.0	35.2	31.9	34.7	26.2	7.0	23.0	2.7	3.4	35.6	5.6	22.5	10.0
	CGG	8.0	4.6	6.8	8.8	17.7	14.8	1.7	15.2	11.7	1.9	1.9	13.0	1.5	7.9	1.2	10.5	16.0	19.0	14.6	16.0	10.4	9.7	2.3	4.8	1.7	0.9	33.0	1.7	6.2	11.9
	CGU	8.5	24.1	15.7	21.4	21.3	6.3	24.2	18.0	9.2	11.4	5.4	8.4	0.9	9.2	23.3	10.6	17.5	6.4	8.2	8.9	10.7	3.9	6.6	10.9	4.9	9.8	6.0	15.2	19.9	16.1
	AGA	2.0	2.1	1.5	6.4	6.4	1.1	3.7	12.6	1.4	10.2	11.2	0.9	27.1	3.8	4.2	1.7	6.6	0.5	3.5	3.3	1.2	0.6	9.7	2.8	13.1	13.6	1.1	7.0	3.1	2.1
	AGG	5.1	1.4	1.3	1.2	5.5	0.6	0.4	6.9	1.5	1.6	35.8	2.6	10.6	3.6	1.2	3.0	5.1	1.7	5.2	5.4	1.3	1.6	4.9	1.4	13.3	2.1	4.4	1.6	2.0	1.4
LEU	CUA	2.1	3.2	2.6	14.0	1.8	2.7	4.9	8.7	3.0	10.1	4.0	4.7	3.7	2.0	9.4	2.5	5.6	0.1	4.5	4.4	1.3	0.6	12.1	1.6	22.0	8.0	0.3	9.1	4.6	10.2
	CUC	18.4	9.9	8.1	7.0	9.0	40.9	5.2	16.7	17.4	7.4	32.8	15.3	5.8	7.6	5.9	19.7	13.4	15.8	28.5	34.1	18.0	37.7	12.1	5.2	7.8	2.0	34.1	7.4	9.9	15.4
	CUG	30.5	54.6	47.4	11.4	58.3	24.6	4.5	34.5	68.1	5.6	18.1	46.4	0.5	24.1	13.5	54.1	49.3	65.0	41.9	34.3	56.5	46.6	16.4	66.0	11.0	2.8	56.8	6.9	49.4	14.6
	CUU	10.7	10.2	5.9	7.1	11.1	5.1	16.8	14.8	6.4	22.1	20.5	5.5	17.5	9.7	11.4	7.2	17.3	12.5	10.8	16.9	10.2	5.4	24.6	4.7	15.8	10.2	1.8	18.2	10.5	10.0
	UUA	1.9	10.9	6.7	25.9	7.5	1.7	46.3	15.9	4.7	27.3	3.0	0.8	37.8	7.2	41.7	2.1	9.5	0.0	1.5	2.2	0.2	0.0	26.0	5.1	34.9	42.2	0.3	20.1	13.4	22.1
	UUG	6.5	11.5	11.6	24.2	5.4	6.2	20.9	12.4	4.6	21.2	1.0	13.2	7.5	31.9	10.7	11.8	15.8	5.2	10.2	10.2	9.2	1.6	15.0	11.2	13.9	11.7	2.6	18.9	11.7	25.3
SER	UCA	4.1	6.8	9.8	6.5	9.9	3.1	13.3	12.4	3.0	18.2	19.1	4.7	23.0	6.2	13.8	3.5	10.0	0.4	4.1	4.4	1.5	1.0	23.2	2.8	16.5	19.0	1.4	16.7	7.4	3.2
	UCC	14.8	9.4	12.5	14.6	14.8	16.1	3.6	7.2	10.7	4.2	8.6	14.4	4.1	14.9	3.5	12.9	11.9	7.4	12.7	14.2	12.7	8.9	12.5	16.8	6.1	2.1	20.7	3.5	10.9	20.9
	UCG	10.0	8.0	11.8	2.5	6.3	18.4	3.5	9.1	11.5	3.9	1.2	19.0	2.0	8.6	3.9	15.1	10.4	24.2	21.0	16.9	19.5	21.8	6.1	12.2	5.3	3.5	15.7	3.0	8.3	2.3
	UCU	6.3	10.4	9.3	21.9	9.2	1.9	20.1	10.7	3.8	16.8	2.6	3.5	6.5	12.8	17.6	3.6	8.7	1.1	3.8	4.4	2.1	1.6	27.1	9.3	9.1	14.1	0.8	16.0	8.7	14.7
	AGC	11.5	15.2	27.8	21.7	14.6	14.1	9.2	13.9	28.1	7.7	9.5	16.1	5.3	16.9	10.4	22.5	15.3	9.5	15.2	14.9	11.7	9.7	12.5	23.6	12.9	8.0	15.1	9.4	16.4	13.9
	AGU	3.1	7.2	10.0	11.3	16.1	4.2	17.5	9.6	3.1	17.7	5.1	2.4	8.5	5.0	15.2	6.6	8.7	0.9	3.9	2.8	1.8	1.0	16.6	3.4	15.6	15.8	1.6	13.9	7.6	8.8
THR	ACA	3.6	6.5	6.3	14.0	11.9	6.3	18.5	12.4	2.3	23.3	22.7	6.0	17.2	7.5	16.5	4.2	13.1	1.1	5.0	5.5	0.8	0.4	22.1	2.8	17.1	31.6	1.8	25.7	6.8	3.7
	ACC	33.6	24.3	36.0	23.2	17.2	24.4	10.8	14.6	32.6	10.9	18.8	42.4	7.5	26.3	16.9	30.8	17.5	31.9	24.3	23.0	36.8	34.2	13.8	38.0	8.2	3.3	40.5	13.1	24.1	31.2
	ACG	10.2	12.7	21.5	1.3	18.2	29.6	9.0	14.3	15.0	10.1	3.2	13.3	4.8	13.3	8.6	12.5	19.9	17.6	17.4	19.1	13.0	17.7	5.5	11.5	17.5	7.7	19.7	7.6	15.5	4.6
	ACU	3.0	10.2	7.5	17.2	8.0	4.2	18.2	11.9	3.6	21.5	3.2	4.6	16.7	9.7	19.2	5.8	9.2	1.8	3.5	5.5	1.3	2.7	24.6	8.1	19.9	18.8	1.2	23.9	8.3	16.3
PRO	CCA	4.1	8.2	5.1	8.0	3.9	4.3	14.4	8.5	3.5	12.9	18.1	6.1	20.3	3.0	11.9	5.2	7.0	0.4	4.3	6.4	1.5	0.4	9.4	4.5	7.4	17.4	1.1	17.5	6.5	2.8
	CCC	32.3	4.3	6.0	10.8	7.9	13.6	2.7	8.3	10.5	2.3	14.1	15.3	2.4	12.8	4.3	12.1	10.2	21.1	12.4	10.6	20.5	15.0	4.1	4.2	7.8	1.4	24.2	2.2	6.0	23.5
	CCG	18.5	23.8	17.5	2.7	22.8	22.2	4.8	7.6	31.5	3.5	5.1	29.7	3.6	15.3	4.4	25.3	16.5	31.8	33.9	23.3	22.6	38.3	4.9	28.4	12.0	4.6	30.9	4.4	23.7	5.6
	CCU	5.8	6.6	5.9	17.2	6.3	2.7	10.9	6.5	4.2	10.4	11.5	4.1	14.8	9.1	14.4	5.8	3.9	1.8	4.6	5.4	2.1	1.4	5.5	4.0	8.9	14.4	1.9	12.3	7.7	11.8
ALA	GCA	11.7	20.8	15.6	37.6	21.3	11.8	37.1	13.9	5.3	24.2	37.4	12.7	48.9	22.5	28.0	12.9	20.9	3.4	11.9	16.0	2.1	5.6	30.7	9.2	19.7	26.4	5.3	28.1	13.5	8.1
	GCC	59.4	23.5	34.4	11.2	35.0	41.7	6.9	22.8	48.4	12.3	25.7	56.2	2.6	44.1	10.3	50.8	38.4	60.6	40.8	46.1	82.0	55.6	13.1	42.3	7.6	4.7	73.6	13.2	26.7	37.0
	GCG	16.9	33.1	32.9	9.2	20.2	40.2	17.4	14.3	46.2	7.9	3.2	37.3	2.9	20.3	14.2	35.1	28.7	63.2	44.9	37.0	26.9	43.7	7.2	39.9	15.8	5.6	47.6	8.5	37.2	10.9
	GCU	9.0	17.4	13.4	59.2	14.9	8.1	28.8	21.1	7.8	29.0	9.9	13.9	30.0	13.0	28.3	13.2	18.0	5.9	10.5	12.8	9.7	9.1	26.3	16.5	14.6	19.7	3.1	37.8	14.7	44.6
GLY	GGA	8.3	7.0	7.6	10.7	10.8	6.3	9.0	13.3	6.7	21.3	21.3	9.2	31.9	8.1	10.0	6.2	11.2	1.6	7.9	9.7	4.0	4.3	17.2	4.5	16.5	18.9	7.9	15.1	7.4	8.8
	GGC	53.1	30.2	44.7	16.1	17.8	41.7	17.0	21.3	45.8	10.5	14.1	57.2	7.7	47.5	18.9	55.0	27.0	58.2	51.8	45.3	60.9	64.9	14.2	59.7	12.7	10.7	60.0	10.9	34.3	23.2
	GGG	9.5	9.7	9.8	4.2	15.7	16.2	8.3	10.6	12.8	6.0	10.9	15.9	7.5	5.1	9.6	9.8	9.7	20.9	9.9	10.7	14.2	15.0	6.4	9.3	16.1	5.7	17.5	6.7	11.3	12.6
	GGU	15.8	27.6	27.3	39.5	20.8	12.3	34.8	15.4	9.9	26.6	28.8	30.0	30.5	22.3	41.0	17.0	17.5	8.1	9.5	12.2	15.7	7.4	21.0	19.3	16.3	26.4	7.6	28.3	19.6	42.1
VAL	GUA	6.0	11.6	8.7	19.9	8.6	3.5	19.1	12.6	5.2	16.0	10.0	4.9	31.9	13.3	13.3	6.4	10.2	0.2	4.1	5.1	0.6	0.2	12.1	6.9	26.2	21.0	2.4	13.7	12.2	14.0
	GUC	27.3	14.2	16.5	6.8	15.9	47.5	4.8	17.6	21.5	10.1	22.5	36.4	2.7	22.6	11.3	25.4	18.0	34.3	33.7	36.6	33.7	37.9	8.8	15.7	13.3	6.1	42.4	12.2	17.7	11.4
	GUG	32.6	25.3	24.0	9.7	29.0	27.0	18.1	18.0	27.7	8.9	20.3	32.8	3.1	17.0	13.7	29.4	31.6	34.1	27.2	21.7	25.1	36.3	5.9	30.6	14.8	7.4	31.9	9.7	24.8	27.0
	GUU	9.8	20.1	12.6	30.9	16.0	7.7	23.0	17.4	8.8	27.5	25.0	11.0	48.4	19.3	30.1	9.1	16.8	6.1	8.7	10.8	12.7	6.4	21.1	12.0	22.6	23.5	1.6	29.4	15.8	21.4
LYS	AAA	11.0	36.5	34.5	23.2	29.8	9.5	58.4	37.4	19.8	58.9	20.3	5.8	72.2	57.4	50.7	10.0	28.4	9.7	7.5	9.0	5.0	6.4	57.0	30.5	33.8	72.2	1.1	55.0	36.7	33.0
	AAG	56.7	12.0	13.7	15.2	14.7	18.4	8.6	27.4	11.1	22.8	37.4	37.2	14.8	12.3	9.1	27.5	24.1	25.0	26.0	26.2	34.0	29.9	13.8	17.6	39.3	16.1	19.9	27.8	12.0	11.6
ASN	AAC	24.2	23.9	33.8	24.4	20.7	27.9	15.1	21.5	23.8	17.4	26.0	32.5	25.4	30.8	18.4	26.6	14.6	18.6	20.3	19.6	19.8	28.4	23.5	37.4	26.4	22.5	21.4	20.2	21.3	22.6
	AAU	8.1	16.3	22.3	15.0	20.2	2.6	40.1	19.3	12.0	37.3	4.2	5.4	15.2	20.1	36.2	8.3	16.0	5.0	9.2	10.7	6.8	3.5	44.1	13.3	24.1	48.3	1.0	39.1	20.2	16.3
GLN	CAA	6.8	13.2	11.6	37.8	5.0	6.3	40.5	19.4	7.8	29.1	1.2	6.7	15.7	23.7	36.3	9.6	10.0	5.3	7.6	7.0	5.6	4.3	30.4	13.1	22.4	33.3	1.6	33.8	12.8	23.5
	CAG	22.2	30.1	30.4	13.2	43.5	27.0	10.0	21.3	44.1	8.1	24.7	26.5	11.2	16.7	10.5	26.6	27.0	21.2	24.3	25.1	27.2	24.3	20.7	33.4	10.1	5.7	24.1	10.6	31.0	11.9
HIS	CAC	22.3	10.7	10.2	3.7	7.4	18.0	6.8	14.4	12.2	5.9	17.3	9.6	11.6	11.6	8.0	14.4	6.8	11.2	13.5	11.8	13.7	16.9	4.2	9.8	3.8	6.0	21.4	5.9	9.4	8.8
	CAU	9.0	11.6	11.7	5.1	12.0	2.1	11.8	18.5	11.0	13.1	6.2	3.2	5.3	8.2	16.9	8.5	11.4	12.0	11.0	9.9	8.3	7.6	12.8	7.8	7.0	18.3	2.4	9.6	11.9	3.5
GLU	GAA	39.1	43.4	24.0	49.4	33.0	29.4	44.1	32.0	26.3	47.5	44.8	14.1	79.6	46.9	40.9	26.2	27.5	28.1	19.3	26.2	24.4	20.6	43.0	25.2	32.5	59.8	9.0	53.9	38.4	46.8
	GAG	20.4	19.2	13.3	10.1	26.9	64.1	11.8	25.6	29.9	11.5	49.0	40.1	6.5	12.5	16.4	29.0	22.8	33.0	38.1	35.3	29.5	27.8	18.5	19.9	23.7	13.1	49.7	16.8	21.7	12.3
ASP	GAC	39.0	21.8	26.2	16.9	24.6	78.4	8.7	22.0	27.6	18.2	43.2	42.4	26.8	28.4	12.3	36.1	17.7	31.0	31.9	34.0	32.8	34.0	17.4	30.8	15.6	13.4	61.1	19.4	21.5	18.1
	GAU	16.8	32.3	33.8	31.7	29.5	14.2	40.4	25.0	27.1	41.5	22.1	11.6	28.3	23.2	40.5	18.8	24.3	22.4	21.4	21.3	17.7	9.9	33.2	27.5	18.2	46.4	3.3	44.8	34.1	28.2
TYR	UAC	23.9	13.4	19.4	27.5	12.6	22.7	8.0	15.7	13.8	13.5	26.3	18.4	17.9	17.5	8.0	18.8	12.4	10.9	13.5	10.9	11.4	21.6	7.0	16.7	23.2	9.1	22.7	15.0	12.6	20.4
	UAU	7.5	15.4	19.4	14.9	16.6	2.2	31.0	18.9	11.8	31.2	4.6	3.1	11.8	18.5	18.0	9.2	15.1	8.9	9.2	10.9	15.0	10.7	18.9	14.5	35.7	35.2	1.2	29.8	17.6	12.5
CYS	UGC	25.3	6.1	4.8	5.4	3.3	2.1	1.9	7.4	10.5	0.9	10.6	5.1	2.6	3.5	3.4	9.7	8.7	12.3	11.5	8.4	7.1	8.0	2.2	6.8	3.6	2.1	9.0	1.0	6.3	2.8
	UGU	7.0	4.7	3.2	3.3	5.8	2.0	6.1	6.5	3.2	4.3	9.5	2.3	7.0	1.3	7.3	2.3	6.1	2.2	2.6	2.0	1.2	0.6	3.6	2.5	4.0	3.8	1.1	2.3	5.0	5.8
PHE	UUC	27.6	18.2	19.1	15.9	16.3	25.5	11.1	17.0	16.9	15.3	24.3	21.0	12.6	17.4	10.3	26.8	19.9	27.0	30.2	27.9	39.6	51.9	10.6	27.8	23.9	11.7	27.9	12.5	15.3	24.2
	UUU	7.4	19.2	16.2	18.7	21.6	2.1	34.4	18.2	18.6	28.5	7.6	4.3	19.3	19.2	26.6	6.8	17.3	6.9	7.7	8.9	5.9	2.5	19.6	8.6	20.3	29.9	0.5	23.9	22.5	24.4
ILE	AUA	4.1	4.1	2.3	2.8	7.5	0.9	7.8	9.8	2.8	12.1	41.4	2.8	15.2	4.5	7.0	2.5	11.7	0.4	3.0	5.1	0.5	0.2	25.8	2.5	45.0	20.1	1.1	5.6	6.7	1.8
	AUC	32.0	26.5	27.5	26.9	21.0	39.0	12.7	22.0	29.9	16.5	19.8	32.0	21.6	25.8	20.4	34.8	22.6	49.0	41.6	42.7	46.7	53.6	16.3	34.5	19.0	11.5	27.7	17.0	24.8	27.7
	AUU	8.3	27.2	22.9	30.5	24.1	3.5	37.3	23.2	17.8	39.3	8.1	9.6	44.6	20.5	38.6	9.8	17.5	4.7	9.3	11.5	6.9	3.3	35.4	13.3	24.9	38.3	0.6	31.0	28.2	36.3
MET	AUG	25.5	26.5	19.9	17.8	31.8	19.8	19.9	22.4	27.2	21.9	28.8	15.9	27.6	20.6	21.9	22.8	25.3	28.5	25.5	26.1	27.7	28.2	20.5	18.8	18.6	20.1	16.4	17.8	25.5	24.4
TRP	UGG	11.8	12.8	15.3	6.1	14.3	8.8	9.3	16.1	16.2	10.1	9.0	13.9	4.4	8.9	11.2	14.0	12.9	13.1	14.3	10.7	15.8	27.2	7.2	12.6	13.1	10.0	15.3	9.8	11.8	13.0
TER	UAA	1.3	2.0	1.3	2.1	1.6	1.6	2.6	2.8	1.2	2.2	1.1	0.9	3.1	1.8	2.2	0.8	2.4	0.8	0.4	0.6	2.0	1.4	1.9	2.0	1.7	2.1	0.2	1.2	2.0	1.9
	UAG	1.7	0.2	0.5	2.2	0.4	0.5	0.2	0.9	0.5	0.6	0.5	1.1	0.3	0.0	0.6	0.3	1.2	0.0	0.7	0.4	0.5	0.0	0.5	0.4	1.3	0.7	0.6	0.4	0.2	1.9
	UGA	0.3	0.8	0.8	0.3	1.5	2.1	0.6	1.1	1.2	0.4	0.9	1.5	0.3	0.8	0.7	1.6	1.9	2.9	2.0	1.7	1.7	3.5	0.8	0.8	4.0	0.4	2.5	0.2	0.8	0.2
	TOTAL	7025	524410	20924	6726	11291	17967	9465	5399	42185	15918	8015	6533	5869	11917	9039	91659	4115	13654	13397	45413	6058	4854	6389	11242	4732	22843	42614	55847	73127	5700

		SYO	TFE	TIP	TRN	TTH	VIB	YEP	ADR	ASV	BTV	FLA	FLB	HIV	HPB	HS1	HS2	HS4	HS5	HS6	HSE	HSV	MCV	MEA	MHV	NDV	NPA	PAF	PIF	PLY	PPH
No.GENES		48	21	63	58	35	71	32	125	29	25	350	72	339	57	71	28	87	60	20	21	27	37	49	25	47	33	20	33	48	60
ARG	CGA	4.1	1.7	11.3	7.2	0.9	5.4	4.7	6.5	6.6	10.1	6.9	4.2	6.8	5.9	6.5	7.2	6.5	8.7	8.1	6.6	3.3	7.8	5.1	3.4	5.3	6.7	4.9	4.5	2.6	8.0
	CGC	21.9	31.4	19.8	22.5	25.5	9.8	11.0	21.2	5.6	8.2	3.8	1.3	2.9	10.1	40.2	42.7	13.4	22.9	4.5	22.1	2.3	12.0	2.7	7.3	6.5	13.3	2.1	1.0	3.4	8.4
	CGG	10.2	15.1	14.7	11.3	34.5	1.4	3.6	8.1	3.0	6.0	6.0	1.4	2.5	8.0	25.0	23.2	14.8	12.9	5.3	9.2	1.7	6.3	6.1	2.7	7.5	2.3	2.4	1.9	2.8	5.9
	CGU	13.0	10.5	12.9	11.9	2.2	16.3	16.4	7.9	8.0	10.8	2.8	3.0	0.6	8.2	5.3	5.7	6.4	14.2	8.2	5.8	2.9	19.9	3.5	8.6	3.2	8.8	2.8	1.0	4.1	8.3
	AGA	0.6	0.8	7.7	9.1	0.7	6.5	5.1	10.1	7.2	21.1	34.2	30.2	37.8	20.5	1.7	1.8	9.4	6.1	13.6	11.0	26.3	11.3	22.5	14.5	11.5	8.3	28.2	29.2	32.0	18.8
	AGG	0.4	1.8	8.3	6.3	14.0	2.1	2.1	8.8	6.2	13.9	19.1	11.6	16.1	11.8	4.6	3.2	15.1	5.6	4.9	9.4	4.7	7.3	19.3	10.8	11.8	4.4	10.7	10.8	19.3	11.5
LEU	CUA	6.3	4.0	8.1	8.6	2.3	14.5	7.4	11.0	12.6	14.1	12.4	17.1	16.3	21.8	5.1	4.3	9.6	8.7	16.1	14.9	15.9	8.9	14.6	11.9	13.2	8.5	17.8	18.0	21.7	12.5
	CUC	24.9	16.8	20.8	12.7	46.4	6.3	7.1	10.3	6.3	5.3	13.9	9.7	9.0	24.4	21.2	21.2	23.6	22.7	10.9	17.0	5.4	17.0	21.2	8.5	16.0	7.3	10.9	9.2	9.9	4.2
	CUG	36.9	43.0	23.3	38.3	37.7	13.2	23.8	28.2	12.3	11.9	16.2	12.3	17.8	21.1	51.8	47.5	39.1	38.5	15.1	23.9	8.0	12.3	25.6	11.4	15.9	13.3	8.2	11.7	17.5	12.7
	CUU	7.4	8.6	21.5	15.3	13.1	15.5	11.8	16.7	26.4	10.9	15.8	14.1	10.3	21.6	5.7	5.3	10.8	7.7	12.1	10.9	20.9	17.8	14.8	17.1	18.6	8.1	21.6	13.0	20.8	10.7
	UUA	4.2	1.7	6.7	18.2	0.6	19.1	21.4	10.0	29.5	27.2	7.3	16.6	21.0	15.2	1.4	1.3	5.0	5.6	27.4	8.0	33.6	18.3	13.3	20.6	19.8	17.0	26.3	25.6	11.5	29.5
	UUG	27.3	12.0	19.0	16.3	8.9	14.8	20.1	18.1	17.6	17.7	12.8	17.1	11.8	24.2	8.2	6.9	10.4	15.1	24.6	14.4	14.0	22.2	18.8	18.7	16.4	34.0	9.4	12.7	17.0	16.8
SER	UCA	3.6	2.5	10.7	10.5	0.1	14.8	17.1	7.5	10.3	13.9	18.6	21.5	11.3	21.3	1.9	1.4	10.6	6.1	11.6	10.4	18.3	17.0	22.1	8.2	21.8	7.6	29.9	35.5	13.1	14.1
	UCC	10.7	14.1	12.9	7.4	16.3	6.0	9.0	14.1	8.5	4.8	9.4	8.5	6.9	23.0	18.6	19.0	18.7	17.9	10.9	17.9	2.7	22.0	14.7	10.4	10.6	7.6	11.6	7.8	12.4	10.3
	UCG	12.6	11.1	14.6	8.8	4.6	9.1	8.4	7.4	4.2	10.7	3.9	2.5	2.1	11.3	15.4	17.9	8.1	17.7	11.3	12.8	2.3	12.2	6.8	3.7	7.1	13.5	3.3	3.7	1.4	4.1
	UCU	9.6	3.2	10.0	8.9	0.7	17.8	17.8	16.0	13.8	6.3	14.6	14.1	7.0	30.3	3.9	3.2	12.1	12.7	18.5	15.9	28.1	27.2	13.1	18.2	26.7	10.0	18.1	17.2	11.6	17.8
	AGC	14.3	19.3	18.2	13.1	12.6	12.0	16.4	17.6	9.7	7.7	15.5	10.0	16.8	5.1	17.5	15.4	18.0	18.9	12.6	18.5	12.0	7.8	17.5	16.7	15.8	18.1	10.6	10.6	11.9	10.8
	AGU	7.7	6.5	9.9	12.0	0.9	14.1	22.2	9.7	12.6	9.7	14.9	12.1	15.0	10.7	2.2	2.4	8.4	8.5	11.0	9.7	16.5	13.9	12.6	20.5	13.3	10.5	14.0	14.1	18.2	16.6
THR	ACA	3.5	3.8	11.0	11.3	0.5	14.9	17.8	10.7	15.6	15.8	28.2	29.7	26.5	15.1	3.9	4.3	12.9	7.2	18.2	15.2	27.0	13.0	21.6	12.5	35.2	11.8	34.7	40.0	23.1	36.7
	ACC	27.7	28.6	17.2	21.5	28.6	13.3	19.7	25.2	11.7	7.0	14.5	15.8	13.1	20.5	30.9	34.5	29.6	26.5	18.5	23.3	3.7	15.2	14.7	13.3	22.3	10.8	12.3	8.8	17.3	14.3
	ACG	11.4	13.1	15.9	12.7	14.4	11.0	13.1	7.3	9.9	17.9	5.3	2.9	3.5	6.7	22.4	25.4	12.9	22.9	16.8	16.8	4.0	10.2	4.3	7.8	4.5	17.2	2.4	5.3	1.7	6.4
	ACU	11.3	3.3	10.7	11.9	0.4	16.5	16.7	16.6	11.9	13.6	17.7	18.1	14.8	22.3	2.6	1.8	10.1	9.0	11.4	16.3	29.1	22.0	15.9	23.5	25.8	15.2	23.7	18.6	23.3	20.8
PRO	CCA	5.6	5.2	12.7	8.2	1.4	15.7	13.5	13.6	11.5	17.5	13.5	18.2	27.5	27.8	7.2	5.6	23.6	7.0	10.0	18.8	22.8	9.6	12.5	16.0	10.8	5.6	22.2	20.8	19.7	23.8
	CCC	16.3	18.1	11.8	7.0	38.5	2.8	5.4	21.7	9.7	4.6	7.1	8.5	9.1	22.2	47.8	47.9	33.7	21.6	10.1	23.8	3.6	13.8	15.1	16.1	11.1	12.2	8.4	6.7	15.2	10.5
	CCG	15.8	18.6	17.1	13.4	13.8	7.8	10.9	13.6	3.2	11.8	4.4	2.9	5.6	10.6	29.1	31.0	15.9	19.0	12.9	16.8	3.0	12.2	8.8	6.0	7.0	10.8	3.6	3.7	3.1	5.0
	CCU	8.4	3.3	10.4	7.8	5.0	12.2	11.7	20.1	14.3	10.1	11.4	14.1	16.6	32.8	4.3	6.3	17.1	10.2	11.4	8.7	27.8	16.9	12.3	17.6	15.7	6.5	14.7	10.3	22.3	22.7
ALA	GCA	21.8	15.3	22.6	17.3	1.8	27.8	27.9	12.4	12.3	16.7	26.7	27.8	31.0	13.3	6.1	4.7	16.5	6.6	11.5	17.7	18.3	11.0	25.2	16.5	31.9	10.5	29.6	22.8	16.7	25.2
	GCC	34.3	48.2	26.7	29.4	76.4	11.8	18.4	26.3	13.2	8.8	12.9	11.6	11.9	14.9	67.8	64.1	44.9	31.9	13.3	35.6	2.9	18.4	17.1	15.9	17.2	15.7	9.9	7.2	14.5	9.8
	GCG	24.9	34.6	25.1	20.6	27.5	19.6	18.9	14.9	6.9	24.3	5.2	2.6	3.7	4.9	42.4	40.9	15.8	19.2	12.9	19.4	3.4	13.0	7.6	7.7	7.8	17.2	2.8	2.5	2.9	4.5
	GCU	31.2	6.8	21.8	17.1	5.2	21.9	27.6	21.4	13.5	18.7	17.1	20.0	16.4	16.5	5.6	6.3	13.2	12.9	15.0	20.3	24.8	33.7	15.9	27.6	16.9	11.4	18.2	13.0	36.6	17.6
GLY	GGA	5.0	7.5	16.8	11.3	5.9	9.0	8.8	17.3	16.6	19.0	31.9	39.1	31.7	24.8	7.1	7.3	17.8	9.5	11.6	16.4	24.5	9.9	21.0	16.4	13.6	6.9	22.7	25.8	22.7	17.0
	GGC	38.8	51.8	24.6	23.2	37.6	18.0	19.4	19.1	10.0	8.1	8.7	8.3	12.2	12.3	34.4	32.5	24.3	25.0	11.0	16.0	6.4	8.4	12.4	19.2	12.0	15.4	5.3	6.4	14.6	12.7
	GGG	10.8	8.0	12.5	10.2	44.3	6.9	19.4	14.2	9.5	11.8	19.2	16.6	19.9	15.9	33.6	30.7	23.5	10.5	6.9	16.5	7.4	11.0	23.8	8.3	24.9	2.4	10.1	11.4	16.8	12.9
	GGU	31.2	15.6	16.8	15.2	3.8	32.6	25.2	14.3	10.5	14.0	10.8	10.4	7.6	8.1	5.3	4.9	12.1	14.0	10.3	11.2	11.2	26.7	16.3	30.4	16.4	9.4	14.7	10.4	16.3	17.5
VAL	GUA	4.8	5.6	8.1	9.7	1.3	18.7	13.0	11.5	13.6	15.5	11.7	16.4	24.5	8.4	3.9	4.3	6.3	10.5	14.6	15.3	20.8	8.3	10.1	10.5	23.5	12.0	14.1	17.2	21.6	21.1
	GUC	23.0	20.6	17.8	16.1	26.6	11.0	12.9	8.6	7.6	8.3	10.8	9.0	8.2	11.2	24.8	29.5	18.8	21.1	15.8	13.4	7.4	18.6	18.2	10.4	17.9	11.2	10.1	9.1	5.0	6.6
	GUG	24.0	34.7	15.9	18.7	60.4	16.9	18.2	28.0	10.0	22.4	19.6	14.6	15.8	14.7	33.4	32.1	28.6	38.2	16.8	23.1	15.1	13.9	16.5	19.0	16.7	26.5	7.8	10.7	15.6	18.4
	GUU	19.1	6.0	17.6	15.5	2.7	22.8	23.3	15.6	13.3	24.6	13.4	12.3	7.8	18.4	5.6	7.8	9.2	9.2	16.3	20.9	23.6	28.0	16.7	28.2	13.8	17.0	15.7	15.6	18.4	16.4
LYS	AAA	20.6	14.3	18.5	37.2	3.3	38.7	37.2	16.9	44.8	28.1	31.9	47.2	32.1	13.2	5.2	3.4	8.2	14.1	39.9	21.1	44.7	25.0	20.3	22.6	18.7	53.8	35.0	42.9	32.0	32.1
	AAG	17.3	36.9	19.3	24.7	44.9	14.2	13.2	18.4	22.5	24.5	20.8	24.6	24.9	6.4	15.3	17.5	20.0	20.1	20.0	18.3	17.5	26.2	27.9	30.2	23.3	21.7	18.9	19.0	15.2	16.3
ASN	AAC	23.7	20.4	17.0	17.1	19.0	26.2	17.5	29.5	20.2	14.3	23.1	27.0	17.1	12.6	22.1	22.0	18.2	24.7	31.2	28.7	19.1	16.5	19.0	16.9	20.4	43.4	21.1	24.8	14.1	13.6
	AAU	9.4	9.0	17.3	21.0	0.8	27.0	37.5	18.9	46.4	28.1	31.4	28.8	28.8	18.8	2.5	2.5	11.8	9.8	28.6	12.7	33.9	18.8	21.7	40.0	29.8	35.2	38.7	42.0	26.1	26.2
GLN	CAA	24.9	9.1	19.9	18.1	2.7	30.4	25.8	14.9	20.2	21.4	22.1	20.4	26.5	20.2	4.9	5.1	12.3	11.7	19.1	11.4	23.6	15.6	22.3	24.4	19.8	25.5	30.0	28.7	27.3	24.9
	CAG	20.5	30.6	20.1	24.7	20.4	13.3	24.5	24.1	15.8	19.3	19.3	7.4	24.9	13.1	23.5	22.7	29.0	25.6	16.1	17.9	11.7	14.4	17.5	25.1	18.3	12.8	14.8	14.7	20.0	21.3
HIS	CAC	11.1	17.1	9.3	10.2	19.6	9.0	6.9	12.6	9.0	9.2	6.3	8.6	11.6	11.3	21.2	21.5	16.5	19.8	11.4	13.3	11.6	9.2	9.8	5.5	6.7	12.7	6.8	6.0	5.9	9.3
	CAU	4.8	9.1	13.1	13.2	1.2	15.3	10.5	8.9	23.1	14.8	9.8	10.7	15.5	11.3	4.5	4.3	9.2	8.0	11.4	6.5	15.8	10.7	9.3	8.2	8.5	10.3	12.6	14.1	14.1	17.1
GLU	GAA	34.0	35.2	32.6	37.0	8.0	40.5	30.3	30.2	36.5	32.9	39.9	47.3	42.4	13.5	9.2	7.9	18.0	21.7	37.6	22.9	42.2	22.0	23.7	23.2	20.7	40.8	33.8	36.0	36.6	33.2
	GAG	20.5	33.6	30.5	26.7	76.6	19.5	21.3	39.1	18.0	32.3	30.7	20.4	31.6	14.6	39.9	42.3	34.3	32.4	18.3	31.9	12.6	26.9	29.8	20.6	19.1	19.2	18.6	17.0	25.7	22.3
ASP	GAC	24.6	38.4	26.8	22.0	43.2	18.6	14.3	28.0	13.9	15.2	20.8	21.5	18.0	18.1	45.6	40.1	28.5	32.5	18.4	30.9	19.6	25.8	24.7	14.2	21.0	34.2	17.0	17.1	20.3	24.5
	GAU	23.0	20.6	31.1	29.7	2.2	39.3	38.4	25.1	32.2	46.1	25.0	29.5	20.8	7.5	8.8	7.8	17.3	16.8	28.2	16.7	27.2	35.3	27.6	34.3	26.5	25.4	36.1	38.9	30.3	34.5
TYR	UAC	19.5	21.8	9.7	13.1	26.5	18.3	8.4	18.6	15.2	13.9	13.1	14.8	10.9	8.9	23.3	26.9	17.9	26.2	16.9	24.1	12.4	17.0	18.3	10.3	18.4	26.5	11.5	11.4	13.1	9.9
	UAU	8.0	9.0	15.3	16.4	2.7	20.8	24.1	11.2	35.4	25.9	14.2	15.2	15.8	15.8	4.7	5.4	9.0	10.1	18.0	10.1	26.4	20.9	12.5	30.4	19.5	23.1	21.0	21.4	21.3	28.7
CYS	UGC	5.0	10.8	8.0	6.0	3.6	4.8	2.8	11.0	12.0	4.6	10.9	13.5	9.5	14.9	12.1	13.7	12.0	13.2	11.3	10.5	7.4	6.6	9.0	10.6	10.6	10.3	7.6	5.4	6.0	10.4
	UGU	2.5	1.7	4.1	4.3	0.3	8.7	3.8	11.2	27.8	6.7	7.2	10.5	13.1	25.8	5.3	4.3	6.2	10.2	16.1	8.1	16.9	13.0	8.2	16.4	11.2	10.4	10.6	10.1	9.4	17.0
PHE	UUC	26.3	23.8	18.1	17.9	24.5	17.4	14.8	12.4	11.3	17.0	20.5	15.1	10.5	26.6	17.7	18.4	15.6	19.2	19.4	9.5	7.1	19.9	15.3	14.6	16.1	12.5	13.0	12.3	9.0	6.4
	UUU	19.2	8.8	18.7	24.7	7.2	28.2	18.9	28.4	32.8	24.4	16.5	19.2	16.6	28.4	17.2	16.1	19.3	20.3	39.6	24.4	40.1	22.4	12.3	33.1	13.1	39.0	18.4	15.1	27.4	32.1
ILE	AUA	0.0	1.8	8.8	11.2	1.3	10.5	11.3	8.9	27.9	24.1	23.5	32.1	27.2	7.5	3.4	3.4	6.8	4.9	21.4	14.1	27.7	7.1	22.4	16.0	24.8	17.3	32.9	43.0	19.5	21.3
	AUC	34.3	40.7	25.0	22.7	34.5	21.3	18.2	11.1	13.9	19.2	19.8	11.3	12.4	18.9	21.5	27.4	17.4	28.3	24.0	17.0	7.3	10.5	30.7	10.2	23.6	14.3	21.3	21.8	3.5	4.3
	AUU	25.8	10.6	20.7	26.4	6.6	29.3	19.6	17.6	39.7	27.4	22.3	20.9	14.1	19.3	5.3	3.6	11.8	9.9	28.2	11.8	32.0	21.7	20.1	25.2	16.0	31.9	31.3	28.2	21.1	22.0
MET	AUG	19.8	28.1	21.7	22.0	19.5	24.0	22.7	25.7	32.1	35.7	39.0	34.3	22.3	19.4	18.3	18.0	19.9	24.6	27.0	21.3	23.6	25.8	25.9	17.0	20.5	24.7	22.4	23.7	24.1	17.5
TRP	UGG	17.3	12.3	9.9	15.2	11.4	12.6	9.9	12.4	20.1	12.5	15.0	11.0	30.2	34.3	11.9	11.9	12.2	14.7	7.5	11.4	10.8	9.6	9.9	12.5	5.3	6.8	9.2	10.9	15.5	13.0
TER	UAA	1.4	1.2	0.8	1.8	0.6	2.4	1.5	1.8	2.3	0.2	1.2	1.1	1.0	1.6	0.5	0.6	0.9	1.2	1.0	1.3	1.9	0.3	0.2	1.6	0.7	2.9	0.8	1.2	2.2	2.1
	UAG	1.5	0.3	0.9	0.7	0.5	0.4	0.1	0.5	1.0	0.8	0.5	1.1	1.2	1.4	0.4	0.9	0.5	0.3	0.1	0.2	0.6	1.1	1.1	0.1	0.3	0.1	0.1	0.2	1.1	0.6
	UGA	0.6	2.0	1.3	1.2	2.0	0.4	1.1	1.0	0.9	0.6	0.8	0.5	1.4	0.8	0.9	0.7	0.5	1.1	1.1	0.3	0.3	4.5	0.6	0.8	0.7	0.1	1.1	0.4	0.4	0.7
	TOTAL	13192	6018	20455	16446	11184	22479	11765	38579	6981	15497	142262	27249	93340	15374	39894	12661	44450	23326	9117	11681	9573	6172	25785	10270	27483	10806	10412	17711	13320	17454

		REO	RSH	SIV	SND	VAC	VAZ	VSV	WHV	F1C	LAM	P22	PMU	PP1	PRD
No.GENES		30	24	91	37	407	71	119	20	22	130	23	49	25	33
ARG	CGA	9.9	2.2	4.3	2.5	5.6	11.9	4.2	3.5	3.5	7.1	10.6	8.1	5.8	2.5
	CGC	7.2	1.7	3.9	1.7	2.4	11.8	1.3	8.8	8.5	15.9	17.9	28.5	16.0	19.0
	CGG	6.2	0.7	2.6	5.7	1.6	11.0	2.7	6.1	0.9	10.2	7.3	12.0	8.1	6.0
	CGU	11.4	2.3	1.4	3.1	8.2	14.6	2.7	8.9	17.2	16.8	13.6	21.7	17.0	10.1
	AGA	13.9	20.5	34.4	24.5	22.0	9.9	20.4	19.6	5.2	9.5	13.3	6.9	7.4	1.0
	AGG	9.5	5.8	17.9	21.0	4.2	5.8	7.9	10.6	1.7	4.7	7.6	5.1	5.5	4.1
LEU	CUA	12.1	18.6	15.3	20.6	18.4	11.1	11.2	10.5	4.4	3.0	7.9	2.6	8.1	5.6
	CUC	10.1	10.1	11.1	16.7	6.2	7.2	12.7	20.6	10.9	9.0	10.8	5.7	9.7	0.8
	CUG	19.2	6.9	14.8	18.7	7.2	12.3	10.4	18.5	15.3	36.3	26.8	44.0	23.1	16.3
	CUU	14.8	12.4	10.0	15.7	11.6	17.6	13.6	21.8	23.7	13.5	18.7	15.4	18.3	19.8
	UUA	16.0	34.6	19.8	16.4	26.5	30.9	16.4	24.1	31.4	8.7	12.2	14.1	15.7	13.3
	UUG	27.1	14.7	17.1	17.5	18.7	17.4	17.8	30.5	15.0	5.7	9.8	8.5	16.3	18.2
SER	UCA	26.7	24.7	13.0	20.4	15.4	13.0	22.3	18.4	16.3	13.4	13.0	15.6	16.0	6.0
	UCC	11.3	12.0	7.9	10.4	10.2	14.5	15.6	23.7	16.1	11.5	6.2	11.7	10.2	10.9
	UCG	14.3	1.2	2.5	7.1	8.6	12.2	4.4	8.0	5.7	8.7	9.2	7.8	5.3	6.1
	UCU	16.9	15.3	9.4	21.5	25.0	14.6	16.8	23.2	50.1	7.4	12.5	5.3	11.0	2.8
	AGC	10.4	14.1	9.5	13.6	6.0	9.0	9.2	6.5	5.7	17.0	13.8	13.7	12.9	16.5
	AGU	13.8	21.9	8.9	11.1	15.4	9.7	13.9	9.1	6.5	11.4	10.6	10.3	10.3	7.8
THR	ACA	18.6	47.8	24.7	26.0	21.2	23.7	24.8	19.4	5.9	13.3	17.1	13.4	21.0	10.1
	ACC	9.8	21.5	11.0	15.8	8.7	19.8	14.6	14.6	11.5	20.8	13.6	17.4	13.7	21.2
	ACG	19.0	0.8	3.9	7.1	9.0	17.3	4.4	8.5	5.7	19.3	9.5	16.7	10.2	13.7
	ACU	24.2	21.8	17.2	17.3	21.9	12.1	18.8	29.5	32.0	8.9	10.8	10.0	13.4	14.9
PRO	CCA	19.4	18.7	27.7	12.1	12.7	17.3	21.2	20.8	7.4	8.9	13.6	10.8	10.7	5.2
	CCC	6.3	9.8	11.1	9.9	4.2	17.8	9.4	17.3	4.6	4.9	1.9	9.4	6.8	9.6
	CCG	8.9	1.0	6.1	6.4	6.3	14.4	8.0	9.3	8.5	17.1	10.3	14.7	7.3	20.7
	CCU	12.4	12.0	13.7	13.5	11.2	11.4	18.6	36.6	23.3	7.4	13.3	10.2	10.7	14.1
ALA	GCA	16.7	22.3	28.1	23.6	11.8	20.7	17.6	12.1	16.3	28.9	32.5	23.3	22.6	15.2
	GCC	12.1	7.3	11.4	12.8	7.0	21.2	8.3	16.4	9.2	28.1	16.8	25.8	17.0	30.7
	GCG	16.0	0.7	6.1	4.5	7.4	18.7	2.1	5.2	7.8	25.4	17.9	23.6	15.2	35.9
	GCU	27.5	14.4	15.9	18.7	13.2	15.0	17.1	21.7	30.5	17.5	24.4	16.0	27.9	21.8
GLY	GGA	22.6	14.7	34.6	17.4	21.3	22.8	35.5	19.4	2.2	12.9	15.2	12.2	11.0	6.4
	GGC	9.1	6.3	13.5	11.4	3.6	9.0	6.3	11.6	25.1	21.3	16.3	19.4	15.3	44.4
	GGG	11.4	4.7	16.9	21.5	2.9	15.4	14.6	9.2	5.2	14.5	9.2	13.1	9.7	11.7
	GGU	17.6	11.5	9.0	16.5	12.7	14.1	12.5	10.9	44.9	19.5	14.9	21.0	19.4	22.1
VAL	GUA	13.9	19.7	20.8	13.1	23.2	19.1	10.8	9.5	15.0	10.0	11.9	11.1	14.4	16.2
	GUC	14.2	8.3	8.8	18.5	8.5	9.1	14.4	11.6	8.5	11.5	10.0	11.8	10.5	4.1
	GUG	28.2	9.8	14.7	20.4	9.1	16.3	15.0	16.0	3.9	24.3	13.6	17.2	12.6	18.3
	GUU	20.0	14.4	8.1	13.1	20.8	23.5	18.7	14.0	49.5	19.0	19.5	23.2	22.6	26.1
LYS	AAA	14.7	62.3	37.4	23.7	48.7	24.5	39.7	26.9	37.3	36.7	42.0	37.1	53.2	36.7
	AAG	18.6	21.7	29.0	28.9	22.7	10.6	26.0	5.6	17.9	19.9	35.2	13.4	27.5	24.7
ASN	AAC	19.6	30.0	16.0	21.8	20.5	19.1	10.9	16.1	10.2	21.2	23.0	18.9	23.4	20.1
	AAU	30.5	48.9	30.0	21.1	47.4	21.7	23.2	29.1	44.4	18.6	19.5	20.2	29.2	30.6
GLN	CAA	19.0	24.1	31.4	13.0	15.5	21.6	16.4	31.3	17.9	9.8	11.9	10.8	19.1	27.0
	CAG	25.4	6.3	28.0	18.8	6.2	13.5	14.1	12.5	21.6	32.9	26.0	30.7	18.1	14.5
HIS	CAC	5.9	7.7	7.7	8.5	6.1	10.8	9.3	10.8	2.8	7.0	7.0	9.1	6.8	4.2
	CAU	8.6	12.8	15.9	10.4	14.5	15.7	18.1	17.7	7.0	11.0	10.6	14.2	11.1	7.3
GLU	GAA	17.5	42.7	42.2	24.9	39.0	32.0	37.6	16.8	19.4	37.4	39.3	42.8	43.1	52.8
	GAG	25.1	10.6	28.1	34.3	14.3	19.9	23.9	8.8	17.0	27.6	33.4	17.8	21.0	3.4
ASP	GAC	23.4	13.1	15.2	21.9	15.3	21.8	25.9	18.9	19.8	24.9	23.0	14.8	13.9	16.8
	GAU	39.1	32.9	24.8	35.8	48.4	32.0	38.9	15.0	35.7	32.5	30.9	30.5	39.6	30.7
TYR	UAC	9.7	11.1	13.0	16.4	15.4	13.0	16.1	10.5	5.9	12.3	17.1	12.4	12.9	11.7
	UAU	17.1	22.2	20.7	19.5	35.6	20.8	23.6	18.6	34.2	17.8	12.7	13.3	19.9	31.4
CYS	UGC	4.2	7.2	13.0	9.3	4.8	6.4	11.1	16.0	4.4	8.9	7.3	9.9	4.7	6.6
	UGU	6.5	10.6	14.3	8.8	15.8	14.9	9.8	13.3	8.1	3.8	4.9	9.0	7.4	2.0
PHE	UUC	13.9	14.1	9.3	17.8	14.1	6.4	23.1	17.8	18.3	14.8	17.9	12.9	18.3	13.4
	UUU	16.7	17.6	17.2	11.1	34.0	33.2	21.0	29.5	36.8	19.8	10.3	22.4	25.2	33.7
ILE	AUA	15.7	46.8	22.7	26.5	37.7	19.0	18.2	17.2	9.8	8.7	11.4	7.4	17.1	8.1
	AUC	21.5	19.2	11.9	28.7	19.0	10.4	23.1	11.4	9.4	20.3	23.6	17.2	22.0	16.1
	AUU	22.2	19.5	13.5	16.5	37.6	26.0	29.6	21.7	35.3	22.5	25.5	30.2	22.8	41.2
MET	AUG	29.6	22.8	20.8	21.6	28.0	20.8	23.3	13.3	17.4	27.6	31.5	25.5	19.9	22.2
TRP	UGG	15.0	6.0	27.6	11.8	7.6	10.9	22.5	34.6	9.6	16.6	15.5	16.0	11.5	10.9
TER	UAA	0.9	1.2	1.0	0.6	2.3	1.3	1.5	0.8	2.8	1.7	2.7	3.9	2.9	3.9
	UAG	0.3	0.4	1.1	0.3	0.6	0.3	0.1	0.0	0.4	0.4	0.0	0.3	1.0	0.3
	UGA	0.9	1.0	1.1	0.3	0.9	0.3	0.5	1.9	1.5	2.6	3.5	2.2	0.2	0.4
	TOTAL	15069	8937	27859	30084	105539	37565	58213	7513	4590	28193	3688	7682	6190	7158

		PT3	PT4	PT7	PZA	BOV MT	CPA CY	CRE CP	DRO MT	EGR CP	MPO CP	MUS MT	MZE CP	MZE MT	PAR MT	PEA CP	RAT MT
No.GENES		49	120	64	23	21	39	29	22	29	52	24	52	22	38	36	32
ARG	CGA	5.5	5.8	7.6	4.2	10.3	12.3	2.3	10.8	7.1	11.3	9.3	14.4	10.7	4.0	9.1	10.2
	CGC	18.3	6.2	15.1	4.0	4.5	5.4	2.4	0.0	3.7	1.9	5.6	6.7	8.6	5.9	3.5	4.7
	CGG	2.2	1.1	2.8	4.0	3.0	1.1	0.0	0.3	0.9	1.1	1.2	4.2	5.7	1.6	3.5	2.3
	CGU	23.8	20.1	22.7	11.3	3.5	35.9	36.0	2.6	17.1	16.1	2.8	14.2	13.6	2.9	14.0	3.2
	AGA	2.9	8.9	5.1	11.6	0.7	11.8	4.9	17.0	19.4	15.4	1.8	19.3	13.1	6.4	11.9	1.2
	AGG	1.2	1.5	3.3	5.7	2.0	0.4	0.6	0.0	4.6	0.9	1.6	5.9	6.5	18.7	4.3	1.1
LEU	CUA	6.9	6.3	10.0	14.6	60.7	4.5	8.7	8.5	6.8	5.2	58.9	13.9	13.1	18.0	14.0	59.5
	CUC	14.6	4.0	10.9	7.6	19.9	0.4	0.4	0.3	0.4	0.8	17.6	6.5	14.7	38.2	6.4	21.5
	CUG	30.2	6.4	22.4	5.9	13.8	0.7	1.3	0.0	0.7	0.8	9.9	5.9	11.7	8.7	6.1	9.4
	CUU	12.5	19.8	16.2	13.4	16.9	15.4	21.3	8.2	18.9	21.5	21.1	19.6	16.8	26.2	21.8	19.7
	UUA	9.3	26.6	11.8	22.6	27.7	73.2	58.8	142.3	50.1	73.1	28.6	35.8	26.9	19.6	36.0	24.7
	UUG	11.4	10.8	10.0	12.2	5.1	3.4	2.4	4.7	16.6	6.2	4.1	20.7	19.1	14.4	21.5	3.6
SER	UCA	8.2	18.3	8.6	17.5	23.0	12.3	19.1	30.3	18.6	14.0	31.7	11.9	17.6	7.9	10.7	26.9
	UCC	12.1	3.2	12.1	4.5	15.8	5.4	1.3	2.4	3.8	3.1	13.2	14.1	12.0	15.8	10.9	16.9
	UCG	3.7	4.0	4.1	4.2	1.3	1.8	2.2	0.8	4.9	1.8	1.4	6.1	9.4	8.3	5.6	1.1
	UCU	15.0	24.9	18.1	14.4	12.9	25.9	19.7	33.3	24.1	25.2	11.8	16.4	14.6	19.4	18.8	12.6
	AGC	9.7	5.5	7.3	9.7	11.4	3.4	4.3	0.3	1.9	1.9	9.0	5.0	8.3	21.3	4.9	11.1
	AGU	5.7	10.6	9.0	19.6	4.1	15.4	11.4	7.7	12.6	16.4	4.9	14.5	18.2	5.2	14.0	4.4
THR	ACA	7.3	17.2	9.6	31.8	37.1	19.4	28.5	22.8	26.0	21.2	37.9	16.4	14.6	9.9	17.9	37.4
	ACC	18.6	6.6	18.5	6.8	24.4	6.2	1.9	0.8	2.1	2.7	21.7	10.6	11.7	18.9	10.9	22.0
	ACG	8.7	4.8	7.7	8.7	4.0	2.2	1.8	0.7	3.5	1.7	2.3	5.9	7.0	11.4	5.6	2.3
	ACU	17.0	27.3	19.7	17.5	14.4	30.8	28.4	28.4	24.1	27.0	15.7	22.7	17.5	12.4	25.2	14.1
PRO	CCA	7.0	13.4	8.7	7.5	23.2	22.3	23.3	10.8	13.5	16.2	32.4	11.2	19.6	4.8	11.7	25.4
	CCC	1.6	1.2	1.6	2.3	16.6	0.5	1.2	1.5	2.5	1.8	9.5	9.5	8.4	15.1	6.8	15.1
	CCG	16.1	5.7	8.2	4.7	0.8	3.6	2.9	0.5	3.1	1.9	0.9	6.1	7.0	0.6	5.7	2.0
	CCU	12.1	14.3	15.1	14.1	12.4	14.7	20.8	19.3	24.4	21.4	8.5	15.5	16.0	7.1	18.9	11.1
ALA	GCA	18.5	20.8	17.5	15.8	24.2	32.4	24.5	8.6	22.8	22.6	23.8	20.1	16.8	12.4	22.2	22.7
	GCC	14.8	5.5	13.8	8.9	24.2	2.9	2.9	3.1	3.5	3.3	24.1	10.4	17.5	21.0	10.1	27.9
	GCG	18.9	6.5	15.1	9.9	2.2	4.5	2.6	1.0	5.0	2.5	2.1	7.9	9.2	4.3	8.7	2.0
	GCU	35.7	33.6	44.3	22.2	18.6	37.5	56.5	31.2	35.9	37.7	16.6	30.6	33.3	20.6	37.2	17.7
GLY	GGA	9.5	19.8	11.3	17.2	25.5	21.9	10.1	38.5	29.0	32.2	28.6	29.1	26.2	11.7	25.9	28.6
	GGC	19.0	8.8	14.3	9.6	17.9	4.7	3.5	0.0	5.0	3.9	12.9	8.4	12.3	18.2	9.6	17.0
	GGG	8.8	3.9	8.7	9.9	11.4	2.7	1.6	5.1	4.0	4.3	8.3	13.9	14.4	11.9	11.6	9.6
	GGU	36.9	31.6	38.2	32.5	8.5	39.3	66.5	13.4	41.0	31.6	10.0	25.0	24.4	7.4	33.9	7.9
VAL	GUA	15.3	19.3	17.5	13.7	21.1	26.6	30.2	22.7	21.4	20.9	19.0	21.5	17.3	12.8	24.5	25.9
	GUC	13.2	5.8	11.7	7.3	14.8	1.3	0.8	0.2	2.4	2.3	11.5	6.7	13.9	17.9	6.8	12.0
	GUG	20.2	5.1	15.0	17.4	8.6	2.7	2.3	1.6	3.8	2.3	7.2	8.5	20.9	9.8	8.7	8.6
	GUU	20.2	32.7	19.6	31.6	10.9	35.9	29.1	20.9	33.4	29.0	13.6	23.4	21.4	19.9	20.0	8.4
LYS	AAA	24.1	63.9	22.3	42.7	21.2	67.8	48.9	23.2	58.1	61.3	26.3	39.6	26.4	27.0	28.2	24.8
	AAG	47.2	16.8	44.4	30.9	11.1	4.9	2.6	1.8	7.4	2.6	7.0	14.4	16.7	41.1	10.7	8.2
ASN	AAC	29.9	15.6	33.4	37.7	24.5	4.9	28.5	2.6	8.4	6.7	25.9	9.0	12.8	32.1	10.2	25.5
	AAU	9.4	41.7	12.0	30.7	14.8	37.3	20.9	52.5	39.8	43.6	16.7	25.7	21.7	18.3	31.5	16.2
GLN	CAA	15.4	22.8	16.9	20.1	17.9	35.9	27.2	18.8	25.1	36.4	18.0	29.2	24.9	7.2	28.9	18.3
	CAG	21.7	11.3	19.3	12.7	6.5	2.4	1.3	1.0	3.1	2.0	4.8	8.9	9.4	10.6	8.3	4.7
HIS	CAC	16.5	4.2	14.1	5.6	15.3	3.1	20.4	2.8	3.4	3.0	16.0	6.1	4.9	14.0	5.0	16.7
	CAU	6.4	12.6	6.3	9.4	8.5	14.0	7.1	19.9	29.1	18.9	11.1	17.7	17.0	6.8	23.3	10.4
GLU	GAA	30.4	59.1	28.9	43.8	22.7	52.6	47.5	19.3	37.7	46.1	25.6	42.6	33.8	9.8	34.8	25.9
	GAG	44.4	10.2	42.4	22.4	9.0	3.3	4.7	0.5	7.4	3.6	7.6	14.4	15.4	25.1	13.2	7.0
ASP	GAC	36.7	15.1	37.9	30.4	14.8	4.4	17.4	2.0	7.2	3.3	13.4	9.6	12.6	19.9	8.0	17.3
	GAU	25.2	47.7	24.3	33.7	9.1	35.0	21.0	11.9	29.3	30.5	12.0	31.2	28.5	11.3	33.4	9.2
TYR	UAC	19.5	9.5	20.5	17.5	20.4	5.6	16.6	6.2	6.4	4.5	14.1	7.3	8.4	29.2	6.0	18.4
	UAU	11.2	31.5	11.9	27.3	15.9	24.8	14.9	36.4	28.1	30.5	17.3	28.7	22.8	19.6	24.6	15.5
CYS	UGC	7.2	3.0	6.2	1.2	4.5	1.1	2.4	0.3	3.4	2.0	5.8	3.0	3.6	11.8	2.6	6.1
	UGU	6.7	6.0	6.2	5.2	2.0	9.6	9.5	8.3	6.1	9.0	3.5	7.8	8.6	4.9	8.2	3.1
PHE	UUC	26.0	10.9	24.9	20.7	32.8	7.3	31.7	6.0	7.1	5.2	32.4	16.7	21.2	49.4	21.1	34.9
	UUU	10.7	31.5	10.2	23.8	26.2	27.9	22.5	91.5	48.0	53.9	28.6	30.1	28.8	76.1	35.9	22.7
ILE	AUA	4.0	11.4	4.5	16.3	48.3	9.6	5.8	61.2	32.7	26.7	48.5	23.4	17.5	12.3	19.8	40.1
	AUC	22.3	11.7	23.7	22.2	37.0	5.8	15.6	6.4	5.9	4.7	34.4	13.8	20.7	25.3	14.6	41.1
	AUU	21.8	51.0	24.8	27.8	40.1	70.7	46.9	92.2	40.1	59.6	55.7	41.6	38.4	21.0	46.8	45.0
MET	AUG	28.3	26.5	30.8	25.2	14.9	21.9	25.0	3.9	25.1	22.7	10.6	22.7	24.0	19.7	20.9	12.0
TRP	UGG	17.9	14.2	16.2	12.5	4.1	5.6	21.9	1.0	18.5	16.5	1.8	15.6	11.0	7.9	22.1	1.9
TER	UAA	3.3	2.8	2.5	1.9	1.0	2.7	2.8	1.8	3.0	3.1	0.5	1.6	1.1	0.2	1.9	2.0
	UAG	0.4	0.3	0.4	0.7	0.0	0.7	0.0	0.2	0.6	0.3	0.0	0.7	1.1	0.4	0.5	0.3
	UGA	1.1	1.1	1.4	1.4	20.1	0.0	0.0	28.1	0.3	0.0	21.5	0.5	0.8	8.1	0.6	22.8
	TOTAL	10358	29243	14751	5758	6031	5517	8315	6129	6764	14960	5674	13500	6177	8999	10302	8948

6		RIC CP	SPI CP	TOB CP	WHT CP	YSC MT
No.GENES		118	50	69	28	50
ARG	CGA	12.3	12.3	13.7	11.7	0.1
	CGC	6.2	5.1	4.6	6.2	0.4
	CGG	5.0	4.4	4.1	2.8	0.8
	CGU	13.7	14.7	18.2	13.4	3.3
	AGA	18.2	14.9	16.6	16.4	24.0
	AGG	8.0	7.3	5.4	6.4	2.3
LEU	CUA	15.1	14.3	13.6	10.4	7.9
	CUC	8.2	5.6	4.5	10.0	0.9
	CUG	6.2	6.0	7.1	7.2	2.5
	CUU	24.5	22.2	19.4	17.8	5.6
	UUA	32.9	32.7	31.5	31.6	100.4
	UUG	20.5	21.7	20.2	16.8	8.3
SER	UCA	12.2	12.2	9.7	8.7	25.4
	UCC	14.4	10.9	11.0	13.8	3.3
	UCG	6.5	6.2	4.1	5.5	1.7
	UCU	19.4	21.0	18.8	14.4	17.4
	AGC	5.1	4.5	5.0	5.5	2.0
	AGU	14.0	14.0	12.6	14.2	12.7
THR	ACA	13.8	14.9	14.4	17.6	21.7
	ACC	10.6	11.6	13.8	11.1	3.5
	ACG	5.7	5.2	5.8	7.9	2.1
	ACU	23.6	21.9	22.4	24.4	16.8
PRO	CCA	11.8	12.9	10.4	12.7	13.2
	CCC	9.3	8.2	7.4	10.2	2.6
	CCG	5.6	5.7	6.9	8.1	1.7
	CCU	18.1	19.6	21.1	19.6	19.0
ALA	GCA	18.7	18.3	21.1	20.4	16.2
	GCC	8.8	11.3	13.2	12.3	3.8
	GCG	7.5	7.9	7.5	8.7	2.4
	GCU	28.5	31.0	37.2	32.7	24.5
GLY	GGA	27.7	28.9	31.5	26.8	12.0
	GGC	8.0	9.3	9.2	13.4	2.1
	GGG	16.0	13.6	13.7	16.1	4.0
	GGU	23.5	28.3	35.0	26.6	38.3
VAL	GUA	22.0	23.0	29.5	24.7	24.9
	GUC	7.2	7.0	6.6	10.2	3.8
	GUG	8.6	8.6	8.5	11.1	2.8
	GUU	23.2	22.6	23.5	23.4	20.5
LYS	AAA	35.7	36.1	32.7	30.4	60.6
	AAG	15.0	12.9	10.0	17.2	10.3
ASN	AAC	11.9	12.4	12.9	9.8	8.7
	AAU	27.6	30.9	29.1	24.2	86.3
GLN	CAA	24.9	27.4	27.0	22.5	19.8
	CAG	8.3	8.1	9.3	11.1	2.4
HIS	CAC	5.8	7.2	6.4	5.5	2.5
	CAU	16.4	19.0	16.4	11.5	17.4
GLU	GAA	39.7	41.0	45.4	42.5	27.7
	GAG	14.8	13.2	13.2	15.5	5.2
ASP	GAC	8.8	9.4	9.5	14.2	6.1
	GAU	28.1	30.7	29.6	31.9	31.3
TYR	UAC	8.2	7.1	8.5	9.1	6.2
	UAU	26.5	24.4	22.5	23.8	47.4
CYS	UGC	3.6	2.7	1.9	3.8	0.6
	UGU	8.7	7.6	6.1	10.4	8.1
PHE	UUC	21.7	19.6	17.3	17.9	17.7
	UUU	35.8	35.4	27.4	28.9	34.8
ILE	AUA	22.1	21.1	17.7	19.1	27.9
	AUC	17.2	15.1	17.4	15.1	11.0
	AUU	38.9	36.8	39.3	40.2	76.5
MET	AUG	24.7	22.0	24.0	24.6	21.6
TRP	UGG	17.9	18.6	14.9	14.9	3.1
TER	UAA	0.1	1.5	2.0	1.9	2.4
	UAG	0.2	0.5	0.5	1.9	0.2
	UGA	0.2	0.5	0.8	1.1	11.7
	TOTAL	23863	17471	19104	5293	14358

Table 2.

***** PRI (Primate genes)

CHP	Chimpanzee
HUM	Human

***** ROD (Rodent genes)

CRU	Chinese hamster
GPI	Guinea pig
HAM	Hamster
MUS	Mouse
RAT	Rat

***** MAM (Mammalian genes other than those in PRI and ROD files)

BOV	Bovine
DOG	Dog
PIG	Pig
RAB	Rabbit
SHP	Sh ep

***** VRT (Genes of Other Vertebrates)

CHK	Chicken
DUK	Duck
ONH	Salmon
SMO	Trout
XEL	Xenopus laevis

***** INV (Invertebrate genes)

APL	Aplysia
BMO	Bombyx mori
CEL	Caenorhabelitis elegans
CHI	Chironomus
DDI	Dictyostelium discoideum
DRO	Drosophila
MOT	Manduca sexta
PFA	Plasmodium
SCM	Schistosoma
SUP	Sea urchin (P.miliaris)
SUS	Sea urchin (S.purpuratus)
TRB	Trypanosoma brucei

***** PLN (Plant genes)

ATH	Arabidopsis
BLY	Barley
BNA	Brassica napus
CRE	Chlamydomonas
EME	Aspergillus nidulans
MZE	Maize
NEU	Neurospora crassa
PEA	Pea
PHV	Bean
POT	Potato
RIC	Rice
SLM	Physarum
SOY	Soybean
SPI	Spinach
TOB	Tobacco
TOM	Tomato
WHT	Wheat
YSA	Yeast (Candida)
YSC	Yeast (S.cerevisiae)
YSK	Yeast (K.lactis)
YSP	Yeast (S.pombe)

***** BCT (Bacterial genes)

ACC	Acinetobacter
AFA	Alcaligenes
ANA	Anabaena
ATU	Agrobacterium
AVI	Azotobacter vinelandii
BAC	Bacillus
BPE	Bordetella
CHT	Chlamydia
CLO	Clostridium
COR	Corynebacterium
DVU	Desulfovibrio
ECO	Escherichia coli
ERW	Erwinia
FDI	Cyanobacterium (F.diplosiphon)
FPL	F plasmid (from E.coli)
HAL	Halobacterium
HEI	Haemophilus influenzae
INS	Insertion element
KPN	Klebsiella
LAC	Lactococcus lactis
MBI	Methanobacterium thermoautotrophicum
MSG	Mycobacterium
MVA	Methanococcus vannielii

NGO	Neisseria
PRM	Prot us
PSE	Pseudomonas
R10	Plasmid R100
RCA	Rhodopseudomonas capsulata
RHB	Bradyrhizobium japonicum
RHM	Rhizobium
RSP	Rhodospirillum rubrum
RSS	Rhodobacter sphaeroides
SHF	Shigella flexneri
SMA	Serratia marcescens
SSP	Sulfolobus SSV1 viruslike particle
STA	Staphylococcus
STM	Streptomyces
STR	Streptococcus
STY	Salmonella typhimurium
SYC	Synechocystis
SYO	Synechococcus
TFE	Thiobacillus
TIP	Agrobacterium Ti plasmid
TRN	Transposon
TTH	Thermus
VIB	Vibrio
YEP	Yersinia

***** VRL (Viral genes)

ADR	Adenovirus
ASV	African swine fever virus
BTV	Bluetongue virus
FLA	Influenza virus A
FLB	Influenza B
HIV	Human immunodeficiency virus
HPB	Hepatitis B virus
HS1	Herpes simplex virus type 1
HS2	Herpes simplex virus type 2
HS4	Epstein-Barr virus
HS5	Cytomegalovirus
HS6	Human herpesvirus type 6
HSE	Equine herpesvirus
HSV	Herpesvirus saimiri
MCV	Cucumber mosaic virus
MEA	Measles virus
MHV	Murine hepatitis virus
NDV	Newcastle disease virus
NPA	Autographa californica nuclear polyhedrosis virus
PAF	Parainfluenza virus
PIF	Human parainfluenza virus
PLY	Polyomavirus
PPH	Human papillomavirus
REO	Reovirus
RSH	Respiratory syncytial virus
SIV	Simian immunodeficiency virus
SND	Sendai virus
VAC	Vaccinia virus
VAZ	Varicella-Zoster virus
VSV	Vesicular stomatitis virus
WHV	Woodchuck hepatitis virus

***** PHG (Phage g nes)

F1C	Bacteriophage f1
LAM	Bacteriophage lambda
P22	Bacteriophage P22
PMU	Bacteriophage Mu
PP1	Bact riophage P1
PRD	Bact riophage PRD1
PT3	Bact riophage T3
PT4	Bacteriophage T4
PT7	Bacteriophage T7
PZA	Bacteriophag PZA (from B.subtilis)

***** ORG (Organelle genes)

BOV	MT	Bovine mitochondrion
CPA	CY	C.paradoxa cyanelle
CRE	CP	Chlamydomonas chloroplast
DRO	MT	Drosophila mitochondrion
EGR	CP	Euglena chloroplast
MPO	CP	Marchantia polymorpha chloroplast
MUS	MT	Mouse mitochondrion
MZE	CP	Maize chloroplast
MZE	MT	Maize mitochondrion
PAR	MT	Paramecium mitochondrion
PEA	CP	Pea chloroplast
RAT	MT	Rat mitochondrion
RIC	CP	Rice chloroplast
SPI	CP	Spinach chloroplast
TOB	CP	Tobacco chloroplast
WHT	CP	Wheat chloroplast
YSC	MT	Saccharomyces cerevisiae mitochondrion

COMPUTER APPLICATIONS IN THE BIOSCIENCES

Volume 5, 1989

IRL PRESS
at
OXFORD UNIVERSITY PRESS
Oxford New York Tokyo

CABIOS COMMUNICATIONS

Vol.5: no.2. 1989
Page 151 – 153

Fast and sensitive multiple sequence alignments on a microcomputer

Desmond G.Higgins* and Paul M.Sharp

Abstract

A strategy is described for the rapid alignment of many long nucleic acid or protein sequences on a microcomputer. The program described can handle up to 100 sequences of 1200 residues each. The approach is based on progressively aligning sequences according to the branching order in an initial phylogenetic tree. The results obtained using the package appear to be as sensitive as those from any other available method.

Introduction

In the recent literature on biological sequence analysis, at least a dozen methods for performing multiple alignments of nucleic acid or protein sequences have been described [e.g. Bains (1986), Sobel and Martinez (1986), Barton and Sternberg (1987), Feng and Doolittle (1987), Santibanez and Rohde (1987), Taylor (1987)]. The motivation for this effort has been the need for the automatic alignment of three or more sequences for the purposes of evolutionary or structural comparisons or for attempting to demonstrate similarity between sets of sequences. In this paper, we describe a strategy which we believe offers the best combination of speed and sensitivity available for any multiple alignment method. We offer a program which can perform multiple alignments of up to 100 sequences of maximum length 1200 residues on a microcomputer in a reasonable amount of time. We judge the program to be 'sensitive' because the results obtained are very difficult to improve by eye.

The strategy we use is essentially that of Feng and Doolittle (1987) adapted for use on microcomputers. The general approach is to progressively align groups of sequences according to the branching order in a hypothetical phylogenetic tree, with gaps that occur in earlier alignments being preserved through later stages. At each alignment stage, a two-sequence alignment algorithm, such as the dynamic programming method of Needleman and Wunsch (1970), is used. For two sequences, the Needleman and Wunsch algorithm gives an alignment that is guaranteed to be optimal for a given set of scoring rules (i.e. weights for aligned residues and penalties for gaps). When this method is used to align two sets of sequences, the score at each position in the alignment is taken from the average score for each residue in one set compared against each residue in the second set. Any gaps introduced into either set of sequences are scored as single gaps. The main difficulty in using this approach on a microcomputer arises from the excessive memory requirements of the Needleman and Wunsch (1970) method—memory usage is proportional to the square of the average sequence length.

In a previous paper (Higgins and Sharp, 1988) we described a strategy for the very rapid multiple alignment of large numbers of sequences on a microcomputer. This method also comprised a progressive approach, using the fast, but approximate, two-sequence alignment method of Wilbur and Lipman (1983). While this approach is extremely rapid and economical with core memory, it works well only for closely related sequences. We did not consider using the exactly optimal method of Needleman and Wunsch (1970) for the progressive alignments because of the excessive memory requirements. However, a recent paper by Myers and Miller (1988) demonstrates how to achieve exactly optimal alignments of two sequences where memory usage varies only linearly with sequence length, without making use of bit packing or secondary disk storage. Thus, a progressive series of alignments of larger and larger groups of sequences, using the method of Myers and Miller (1988) for each alignment, is the key to the current approach.

System

The program described in this paper was written in standard FORTRAN 77 and compiled using the Microsoft FORTRAN compiler, version 4.0. Program performance was tested on an IBM AT compatible microcomputer, running at 10 MHz with no maths coprocessor, 640 kbytes of memory and a hard disk. This program (CLUSTAL4) is an extension to the package described in Higgins and Sharp (1988). Copies of the executable files, documentation and test data files will be sent on request. Please send three 5.25 inch floppies formatted to 360 kbytes, or one high density 5.25 inch floppy formatted to 1.2 Mbytes.

Algorithms

The program takes, as input, a dendrogram produced by applying the UPGMA method (Sneath and Sokal, 1973) to a matrix of similarity scores between all pairs of sequence to be aligned. The similarity scores are calculated as the number of exactly matched residues in a Wilbur and Lipman (1983) align-

Department of Genetics, Trinity College, Dublin 2, Ireland
**To whom correspondence should be addressed*

ment between two sequences, minus a fixed penalty for every gap. For short sequences, several similarity scores per second can be calculated on a microcomputer using the package described in Higgins and Sharp (1988).

The sequences are then aligned in groups corresponding to the branching order in the dendrogram. The alignments are carried out using the method of Myers and Miller (1988), adapted for use in a multiple alignment context. Myers and Miller took the distance minimizing algorithm of Gotoh (1982) and applied a 'divide and conquer' strategy (attributed to Hirschberg, 1975) to give alignments in linear space; as a consequence memory usage is linearly related to sequence length. Briefly, the method is based on finding the optimal mid-point of an alignment. When this is found, the matched symbols (two aligned residues, or one residue opposite a gap) are part of the final alignment. The rest of the alignment is found by recursively finding optimal mid-points on either side of the initial mid-point. In this context, the optimal mid-point can be defined as the aligned symbols at the centre of the optimal alignment. The centre is taken to be half way along one sequence.

Two modifications were needed to adapt the Myers and Miller algorithm for our program. Firstly, all real number operations were converted to using 2-byte integers. On a 16-bit microcomputer without a maths chip, this increases the speed of each alignment by a factor of 30. Indeed the speed approaches that described in Myers and Miller for their program running on a VAX 11/780. Secondly, the scoring system was modified to allow all residues at a given position in each group of sequences to contribute to the alignment scores. For proteins, we use the log-odds amino acid similarity matrix of Dayhoff (1978) to score aligned residues. The similarity matrix was rescaled to give positive integer weights between 0 and 25 and then converted to a difference matrix by subtracting each value from 25. Thus two aligned tryptophans have the lowest distance (0) while a cysteine aligned with a tryptophan has the largest distance (25). For nucleic acid sequences we use a three-tier weighting system where identical residues have zero distance, transitions have a distance of 5 and transversions have a distance of 10. A variable gap penalty is used; a fixed penalty is added to the alignment distance for every gap and an extra penalty is added for every item in the gap. Gaps that are introduced into a pre-aligned group of sequences are scored as single gaps. Both of these penalties can be specified at run time.

In order to calculate the alignment scores between two clusters of sequences, the gaps that are already inserted in the two clusters (from earlier alignment stages) are treated as being fixed. Thus, each cluster may be thought of as a single sequence,

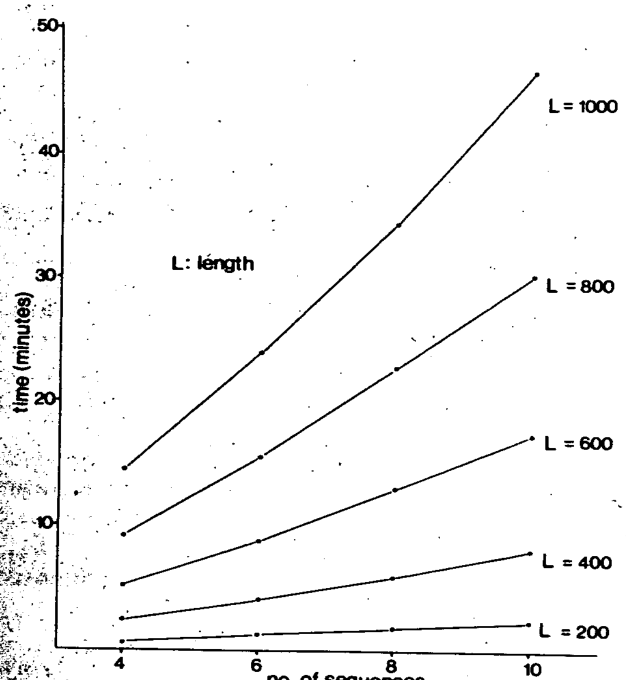


Fig. 1. Times required for the multiple alignment of different numbers of sequences of different lengths. Each curve represents the times for truncated fragments of a given length, L; this example used the HIV pol protein. Times for calculating the similarity matrices or dendrograms (maximum of <3 min) are not included.

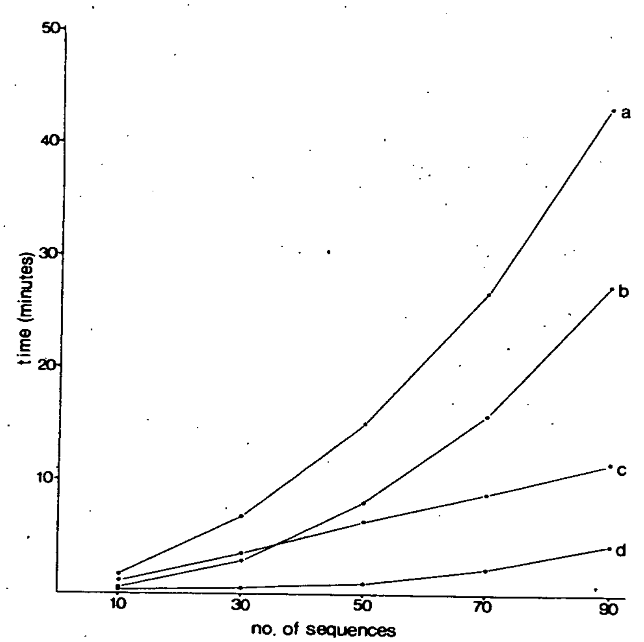


Fig. 2. Times for aligning different numbers of sequences. This example used globin sequences (alphaglobins, betaglobins and myoglobins) each truncated to 140 residues. The four curves show times for different parts of the multiple alignment process: (a) total time, including calculation of similarity matrix and dendrogram, (b) calculation of similarity matrix, (c) multiple alignment, (d) construction of UPGMA dendrogram.

where the residues at each position are the 'average' residues at that position in each of the sequences represented by the cluster. In order to calculate the weight between a position in one cluster and another position in the second cluster, one takes the arithmetic average of all the pairwise weights between each residue in one cluster versus all of those in the second. For two clusters with K and L sequences each, this involves taking the average of K times L weights at each alignment position. This becomes very time-consuming with many sequences, but can be speeded up by pre-calculating the weight of each position in each cluster versus each possible residue.

Results and discussion

The speed of the program can be demonstrated by aligning different numbers of sequences of different sizes. We find that the speed is almost totally independent of the characteristics of the sequences, apart from length. This was also noted by Myers and Miller (1988) for their two-sequence alignment program. Figure 1 shows the times required for a series of multiple alignments of sequences from 200 to 1000 amino acids in length. One expects the alignment times to vary with the square of the sequence lengths and a visual inspection of the figure confirms this. The time required to align different numbers of sequences varies approximately linearly. A slight departure from linearity is evident with the longer sequences. This confirms the effectiveness of our stategy of pre-calculating the weights at different positions in each cluster. The times required for calculating the initial similarity matrices and construction of the dendrograms are not shown. These only need to be calculated once for any multiple alignment. The slowest dendrogram to construct was that for the ten 1000 residue sequences. This took under 3 min.

Figure 2 shows the times required to align from 10 to 90 sequences of 140 amino acids each. In this case the times for each of the various calculations are shown. The similarity matrices and dendrograms were constructed using the programs CLUSTAL1 and CLUSTAL2 (Higgins and Sharp, 1988) respectively. For large numbers of sequences, the calculation of the initial similarity matrix is the dominant time-consuming factor. For 90 sequences, this requires the calculation of 4005 values. Nonetheless, the times involved are quite practical on a microcomputer.

The sensitivity of the program is more difficult to demonstrate. Our basic criterion in determining sensitivity is to assess the ease with which the resulting alignments can be improved by manual adjustment. By this criterion, we find the results of our program to be excellent. In this respect, the program can be used confidently to replace the usual manual alignment of sets of closely related sequences for publication. Of greater scientific importance is the usefulness of the program for aligning regions of homologous secondary structure or in reconstructing evolutionary events between distantly related sequences. This is more difficult to demonstrate. Barton and Sternberg (1987), Feng and Doolittle (1987) and Taylor (1987) discuss these questions in detail. It is possible that no single method will be ideal for these purposes. As a general observation, we find the alignments produced by our program to be at least as good as those produced by each of the above authors.

Acknowledgements

This is a publication from the Irish National Centre for Bioinformatics. This work was supported by grant number BAP-0137-IRL from the European Community Biotechnology Action Programme. We are particularly grateful to Gene Myers for sending us the source code of his linear space alignment program.

References

Bains,W. (1986) MULTAN: a program to align multiple DNA sequences. *Nucleic Acids Res.*, **14**, 159–177.
Barton,G.J. and Sternberg,M.J.E. (1987) A strategy for the rapid multiple alignment of protein sequences. *J. Mol. Biol.*, **198**, 327–337.
Dayhoff,M.O. (1978) A model of evolutionary change in proteins. Matrices for detecting distant relationships. In Dayhoff,M.O. (ed.), *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Washington DC, Vol: 5 Suppl. 3, pp. 345–358.
Feng,D.-F. and Doolittle,R.F. (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.*, **25**, 351–360.
Gotoh,O. (1982) An improved algorithm for matching biological sequences. *J. Mol. Biol.*, **162**, 705–708.
Higgins,D.G. and Sharp,P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, **73**, 237–244.
Hirschberg,D.S. (1975) A linear space algorithm for computing longest common subsequences. *Commun. Assoc. Comput. Mach.*, **18**, 341–343.
Myers,E.W. and Miller,W. (1988) Optimal alignments in linear space. *CABIOS*, **4**, 11–17.
Needleman,S.B. and Wunsch,C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, **48**, 443–453.
Santibanez,M. and Rohde,K. (1987) A multiple alignment program for protein sequences. *CABIOS*, **3**, 111–114.
Sneath,P.H.A. and Sokal,R.R. (1973) *Numerical Taxonomy*. Freeman, San Francisco.
Sobel,E. and Martinez,H.M. (1986) A multiple sequence alignment program. *Nucleic Acids Res.*, **14**, 363–374.
Taylor,W.R. (1987) Multiple sequence alignment by a pairwise algorithm. *CABIOS*, **3**, 81–87.
Wilbur,W.J. and Lipman,D.J. (1983) Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA*, **80**, 726–730.

Received on October 7, 1988; accepted on December 6, 1988

Journal of Molecular Evolution

Volume 25 (1987)

Springer International

J Mol Evol (1987) 25:351–360

Journal of
Molecular Evolution

Progressive Sequence Alignment as a Prerequisite to Correct Phylogenetic Trees

Da-Fei Feng and Russell F. Doolittle

Department of Chemistry, University of California–San Diego, La Jolla, California 92093, USA

Summary. A progressive alignment method is described that utilizes the Needleman and Wunsch pairwise alignment algorithm iteratively to achieve the multiple alignment of a set of protein sequences and to construct an evolutionary tree depicting their relationship. The sequences are assumed a priori to share a common ancestor, and the trees are constructed from difference matrices derived directly from the multiple alignment. The thrust of the method involves putting more trust in the comparison of recently diverged sequences than in those evolved in the distant past. In particular, this rule is followed: "once a gap, always a gap." The method has been applied to three sets of protein sequences: 7 superoxide dismutases, 11 globins, and 9 tyrosine kinase-like sequences. Multiple alignments and phylogenetic trees for these sets of sequences were determined and compared with trees derived by conventional pairwise treatments. In several instances, the progressive method led to trees that appeared to be more in line with biological expectations than were trees obtained by more commonly used methods.

Key words: Multiple sequence alignments — Evolutionary trees

Introduction

The evolutionary relationships of sets of protein (or nucleic acid) sequences are commonly depicted in the form of trees (Fitch and Margoliash 1967; Dayhoff et al. 1972; Moore et al. 1973; Sankoff et al. 1982; inter alia). Indeed, the digital nature of sequence data makes them more amenable to such treatment than is the case with many more qualitative biological characters. Most current schemes for constructing trees from sequences use a simple difference matrix, the elements of which are assembled by performing pairwise comparisons of all the sequences under study (Fitch and Margoliash 1967). A topology is found by classifying the sequences according to their differences, which ought to be a reflection of the evolutionary distances among them. For the most part, the principle of parsimony is rigorously adhered to, and the best trees are thought to be those that can account for the extant sequences by the smallest number of genetic events. The two important features of a tree are its topology, or branching order, and its branch lengths, which ought to be proportional to the true evolutionary distances.

In principle, the construction of an evolutionary tree based on sequence data ought to be a simple matter: all one has to do is cluster the sequences according to their similarities. In practice, uncertainties and ambiguities concerning both the topology and branch lengths are common, and enormous effort is often expended in finding the "best tree" (e.g., Fitch 1977; Penny and Hendy 1986). Finding the correct tree should depend on assembling a matrix that best describes the differences among the sequences, and this depends, in turn, on properly aligning the sequences (Hogeweg and Hesper 1984). The alignments can be obtained either by schemes that maximize similarity (Needleman and Wunsch 1970) or with those that minimize differences (Sellers 1974). If a similarity scheme is used, the scores must be transformed appropriately into measures of distance.

Offprint requests to: Da-Fei Feng

Ordinarily, alignments of either type are performed pairwise. The problem is that when the various paired alignments are grouped, they are seldom consistent one to another. Thus, when sequence A is paired with sequence B, gaps may appear at various locations, but when either A or B is aligned with a third sequence, C, the arrangement of gaps may be entirely different. Heretofore, this problem has been circumvented by making a multiple alignment of all the sequences by the judicious shifting of the sequences as needed to minimize differences ("eyeball" alignment).

The flaw in the approach is that these multiple alignments have, like pairwise alignment schemes before them, been subject to rigorous attempts at parsimony. Obviously, the closer two sequences resemble each other, the more confidence one has in the alignment. But in most multiple alignment schemes where maximum parsimony is sought, no distinction is made with regard to the confidence one has in a particular pairwise alignment. It seems to us folly that a gap should be discarded in an alignment of two closely related sequences merely because an alignment with some distantly related sequence might be improved.

To this end, we have devised a scheme of progressive sequence alignment that has a higher intrinsic regard for recent events than for distant ones. It is still based on a maximization of similarities, but it follows the simple rule "once a gap, always a gap." It is able to accomplish this by inserting neutral elements into sequences once gaps have been established. The sequences are aligned progressively, beginning with the most similar pair and continuing with the addition of the next most similar sequence or set of sequences. The difference scores obtained from the final alignment of all sequences are then used to construct the evolutionary tree. Ambiguities may still arise, of course, since the preliminary matrix of similarities (or differences) based on pairwise comparisons will often include what we call "better but less reliable" scores. These can be sorted out by testing alternative trees. Because it is impractical to consider all possible pairwise orders, we have adopted an effective compromise whereby reasonable alternative arrangements are explored progressively.

In this paper we describe the details of the method and apply it to several groups of protein sequences. Trees constructed by this approach can differ significantly from those assembled by traditional schemes, but they are often in accord with what might be expected on the basis of organismic phylogenies. The method has the added virtue of providing multiple sequence alignments quickly and simply by completely objective criteria.

Methods

Studies were performed on a DEC 11/730 VAX computer with the UNIX (Berkeley 43) operating system. The plotting package for use with a Nicolet Zeta plotter was written by Steve Dempsey of the U.C.S.D. Chemistry Department Computer Center. All utility programs were written in the C programming language (Kernighan and Ritchie 1978). The ensemble of programs dealing with sequence alignment and tree building can be contained by sending a blank magnetic tape to the authors.

Definitions. For purposes of description only, we would like to distinguish between *simple* and *compound* trees. Simple trees are those in which the branching order follows the simple clustering (((AB)C)D) etc., whereas compound trees have subclusters, as in ((AB)(CD)E). *Neutral elements* are simply characters (Xs) that are filled into sequences when gaps occur. They are neutral in the sense that they are invisible to the scoring system used to establish subsequent alignments, which is to say when X is matched with any other residue, the value is equal to zero. *Negative segments* are those internodal connecting distances with negative values that occasionally emerge from Fitch-Margoliash trees when data scatter confounds the segment averaging (or least-squares treatment). *Percent identity* is taken as the number of identities per 100 aligned residues.

Sequences. Amino acid sequences were taken from an updated version of the NEWAT database (Doolittle 1981). Primary references to the nine tyrosine kinase sequences and nine of the globin sequences have been provided in an earlier study (Feng et al. 1985). The additional globins used in the present study are from lamprey (Zelenik et al. 1979) and the bacterium *Vitreoscilla* (Wakabayashi et al. 1986). The superoxide dismutase sequences studied are human (Jabusch et al. 1980), bovine (Steinman et al. 1974), swordfish (Rocha et al. 1984), fruitfly (Lee et al. 1985), maize (Cannon et al. 1987), yeast (Johansen et al. 1979), and photobacter (Steffens et al. 1983).

Pairwise Alignments. The algorithm of Needleman and Wunsch (1970) was used in a three-matrix form (Fredman 1984) and utilized the Mutation Matrix of Dayhoff et al. (1978) in its scoring. The algorithm was actually employed in several slightly different settings. In the first, a program called SCORE aligns pairs of sequences in the conventional way and stores their alignment scores in a table. The similarity scores obtained from the alignments are converted to difference scores by the relationship

$$D = -\ln S_{eff} \times 100 = -\ln \frac{S_{real} - S_{rand}}{S_{ident} - S_{rand}} \times 100$$

where S_{real} is the alignment score itself, S_{rand} is the score obtained with random sequences of the same lengths and compositions, and S_{ident} is the average score of the two sequences being compared when each is aligned with itself. In practice, in these initial pairwise comparisons we use an average value for S_{rand} based on many previous observations (Feng et al. 1985). Inasmuch as this initial set of comparisons is assumed to be imperfect, no precision is lost by the modification, and considerable time is saved by the omission of numerous jumble comparisons. The value used, after normalization to a standard length, was 770, the average random score for numerous comparisons of many different kinds of sequences (Feng et al. 1985).

The Needleman-Wunsch algorithm is used in a second series of alignments in a mode in which gaps are concurrently filled with neutral elements. In the main version, DFalign, sequences are aligned successively. Should the tree in question be a com-

pound tree, subclusters are first prealigned with a simpler version of the program called PREalign.

Tree Building. A program based directly on the Fitch and Margoliash (1967) procedure was written in our laboratory by Mark Johnson. The program, BORD, was used to establish preliminary branching orders. Simply put, the smallest difference score is identified and a new matrix constructed that contains the average distances between members of the first pair and remaining members of the set. The procedure is repeated until all scores have been incorporated. A second program, BLEN, was used for determining branch lengths of the final tree. This program employs a least-squares approach as described by Klotz and Blanken (1981). In the event that a tree contains one or more "negative segments," the "nearest alternative" trees are considered and their scores compared. *Nearest alternative* trees are those in which the branches immediately adjacent to a negative segment are switched. The program TREEplot, also written by Mark Johnson, puts the data in an appropriate form for the Zeta plotter in order that dendrograms can be issued directly.

Outline of the Progressive Method

Pairwise Alignments

For n sequences, the number of pairwise alignments required for the initial matrix amounts to $(n - 1) \times n/2$. To this end, a simple UNIX shell program was constructed for running each comparison serially with the program SCORE; the resulting difference scores are automatically stored in a suitable file.

Identification of Most Closely Related Pair

The program BORD takes the output from SCORE and establishes a preliminary order of the sequences. The program BLEN uses the difference matrix from the SCORE program combined with a simple "connectivity table" to give branch lengths; the connectivity table merely puts all the connecting segments in tabular form. BLEN is only used at this point if trees based on pairwise comparisons are going to be prepared. The BORD program reveals whether or not the starting tree is simple or compound. In the case of compound trees, subclusters are prealigned with the program PREalign, which aligns the cluster and fills the gaps with neutral elements (Xs).

Progressive Insertion of Neutral Elements

The program DFalign, which is the heart of the procedure, is used to generate the multiple alignment. It begins by inserting neutral elements (Xs) in any gaps that occur in the aligned pair with the highest similarity score. After the original pair has been established and the gaps fixed, the next nearest relative or set of relatives is brought in and a new alignment made and a score determined. The key to this alignment is that new gaps can be incorporated into either sequence, but the earlier gaps are preserved. The first ternary arrangement, ABC, is then compared with the alternative BAC, the higher score being used to set the path for the next alignment. Similarly, when the next sequence is brought in, the arrangement ABCD is scored and compared with ABDC. Prealigned subclusters are maintained as separate units, however. The procedure is continued until all sequences have been incorporated.

Scoring the Final Alignment

The final alignment is scored with a modified regimen that recognizes the fixed nature of the gaps. Moreover, because the gaps are fixed, it is unnecessary to use an alignment program at this stage. Instead, a scoring system is used that measures S_{real} and S_{ident} in the usual way, but that employs a program, SHUFFLE, for determining S_{rand}. SHUFFLE randomizes each sequence numerous times while holding the gaps constant.

Constructing the Tree

The program BORD is used to obtain the new branching order and the program BLEN to determine the branch lengths. If any negative segments result, alternative trees with the branches on either side of the negative segment reversed are constructed and a new set of branch lengths calculated. If negative segments are still present, the alternating procedure is continued until they disappear, although we have not yet encountered a situation where more than one switch was necessary. The program TREEplot is used to produce the final dendrogram. A schematic outine of the programs called from start to finish is present in Fig. 1.

Results

Superoxide Dismutase

The sequences of seven copper–zinc superoxide dismutases—human, bovine, swordfish, fruitfly, maize, yeast, and photobacter—were subjected to a conventional pairwise alignment scheme and a tree constructed by the Fitch and Margoliash (1967) procedure (Fig. 2a). The same seven sequences were then treated by the progressive procedure and a tree generated (Fig. 2b). The trees differ both in branch order and branch length. More to the point, the progressive procedure yields a tree that corresponds to the accepted phylogeny of the organisms, whereas the conventionally generated tree does not.

In fact, the initial tree issued from the ordinary Fitch and Margoliash (1967) treatment had the expected phylogenetic branching order, but contained

a negative segment. When the nearest alternative tree was examined, generated by reversing the branches on either side of the negative segment, the sum of branch lengths was lowered, and a "better tree" with no negative segments emerged (Fig. 2a). The tree contradicts what is known of the evolutionary relationships of the organisms involved, however, in that the branch to the yeast sequence comes off above the branch to the *Drosophila* sequence.

It should be emphasized that the multiple alignment (Fig. 3) used to obtain the final tree was obtained by strictly objective criteria and without recourse to "eyeball" manipulation. Moreover, the overall similarities, as reflected in the percent identities, are more in line with the true distances separating the organisms than are those observed in the original pairwise alignments (Table 1).

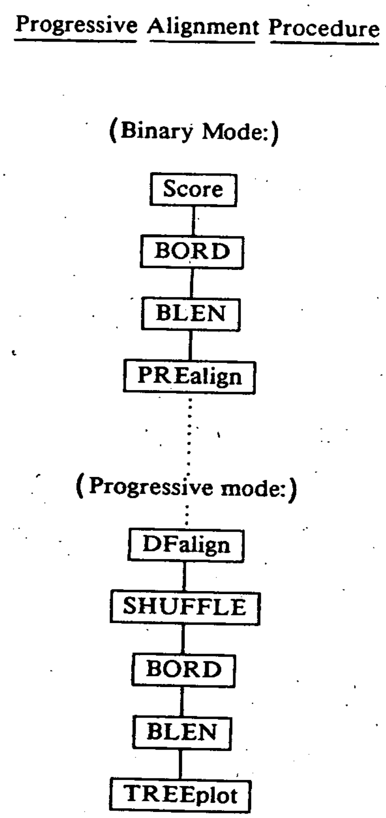


Fig. 1. Flow chart of progressive alignment procedure. Program names are shown in boxes. The program BLEN in upper portion of figure may be omitted if a tree based on pairwise alignments is not going to be constructed.

Hemoglobins

Eleven different globin sequences covering a broad spectrum of types were subjected to pairwise alignments and an initial tree constructed from the resulting difference matrix (Fig. 4a). The tree was similar to those presented in previous reports in that cyclostome globins (hagfish and lamprey) branch off in advance of the myoglobin–hemoglobin α-chain divergence (Goodman et al. 1974; Hunt et al. 1978; Feng et al. 1985). When the same 11 sequences were subjected to the progressive alignment procedure, the tree that emerged reversed the order to the more biologically reasonable situation in which the cyclostome globins are clustered with those of other vertebrates (Fig. 4b).

Also of interest are the relative positions of the plant and invertebrate hemoglobins. In the tree obtained from pairwise alignments, the plant and bacterial hemoglobins appear to be more closely related

Table 1. Percent identities calculated from binary (upper triangle) and progressive (lower triangle) alignment methods

	Superoxide dismutases						
	sodh	sodb	sodl	sddm	sdmz	sods	sdpb
sodh		82	67	60	62	53	31
sodb	82		74	57	61	55	35
sodl	67	72		59	59	56	35
sddm	59	59	58		68	54	31
sdmz	60	60	58	68		57	32
sods	51	52	54	51	54		30
sdpb	31	35	34	31	34	31	

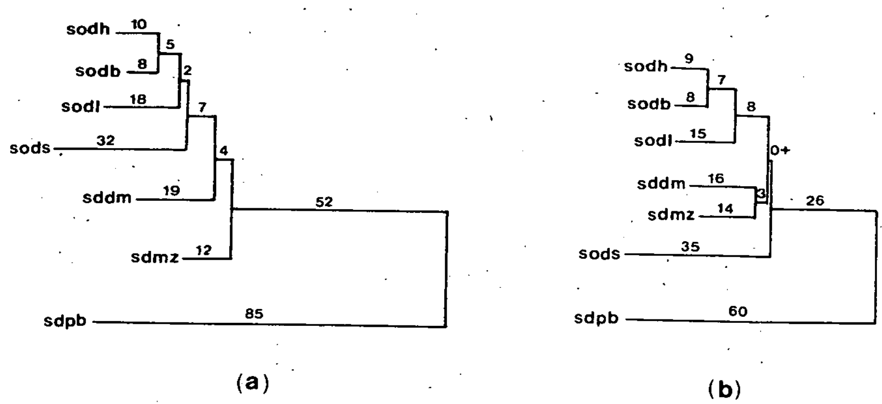


Fig. 2. Phylogenetic trees for seven superoxide dismutases, as determined by **a** simple pairwise alignments and **b** progressive multiple alignment. The four-letter designations are sodh, human; sodb, bovine; sodl, swordfish; sddm, fruitfly; sods, yeast; sdmz, maize; sdpb, photobacter. The same designations are used in Fig. 3 and Table 1.

```
                       *        *                                       ***  *   *          *            *   *   *
sodh         ATKAVCVLKGDGPVQGSINFEQKESDGPVKVWGSIKGLTE GLHGFHVHQFG    NDTAGCT     SAGPHFNP LSRK
sodb         ATKAVCVLKGDGPVQGTIHFEAK  GDTVVVTGSITGLTE GDHGFHVHQFG    DNTQGCT     SAGPHFNP LSKK
sodl         VLKAVCVLRGAGETTGTVYFEQEGNANAVGKGIILKGLTP GEHGFHVHGFG    DNTNGCI     SAGPHFNP ASKK
sddm         VVKAVCVING   DAKGTVFFEQESSGTPVKVSGEVCGLAK GLHGFHVHEFG    DNTNGCM     SSGPHFNP YGKE
sdmz         MVKAVAVLAGT DVKGTIFFSQEGDG PTTVTGSISGLKP GLHGFHVHALG    DTTNGCM     STGPHFNP VGKE
sods       VQAVAVLKGDAG   VSGVVKFEQASESEPTTVSYEIAGNSPNAERGFHIHEFG    DATNGCV     SAGPHFNP FKKT
sdpb      QDLTVKMTDLQTG  KPVGTIELSQNKYG  VVFTPELADLTP GMHGFHIHQNGSCASSEKDGKVVLGGAAGGHYDPEHTNK

          *   *      *  **                          *          *       *    *                 *   *  *   ** *
sodh      HGGPKDEERHVGDLGNVTADKDGVADVVIEDSVISLSGDHCIIGRTLVVHEKADDLGKGGNEESTKTGNAGSRLAcGVIGIAQ
sodb      HGGPKDEERHVGDLGNVTADKNGVAIVDIVDPLISLSGEYSIIGRTMVVHEKPDDLGRGGNEESTKTGNAGSRLAcGVIGIAK
sodl      HAGPKDEDRHVGDLGNVTADANGVAKIDITDK ISLTGPYSIIGRTMVIHEKADDLGRGGNEESLKTGNAGSRLAcGVIGTE
sddm      HGAPVDENRHLGDLGNIEATGDCPTKVNITDSKITLFGADSIIGRTVVVHADADDLGQGGHELSKSTGNAGARIGcGVIGIAK
sdmz      HGAPEDEDRHAGDLGNVTAGEDGVVNVNITDSQIPLAGPHSIIGRAVVVHADPDDLGKGGHELSKSTGNAGGRVAcGIIGLQG
sods      HGAPTDEVRHVGDMGNVKTDENGVAKGSFKDSLIKLIGPTSVVGRSVVIHAGQDDLGKGDTEESLKTGNAGPRPAcGVIGLTN
sdpb      HGFPWTDDNHKGDLPALFVSANGLATNPVLAPRLTL    KELKGHAIMIHAGGDNHS     DMPKALGGGARVAcGVIQ
```

Fig. 3. Multiple alignment of seven superoxide dismutases determined by progressive method. Asterisks denote locations where all seven residues are identical. See legend to Fig. 2 for four-letter designations.

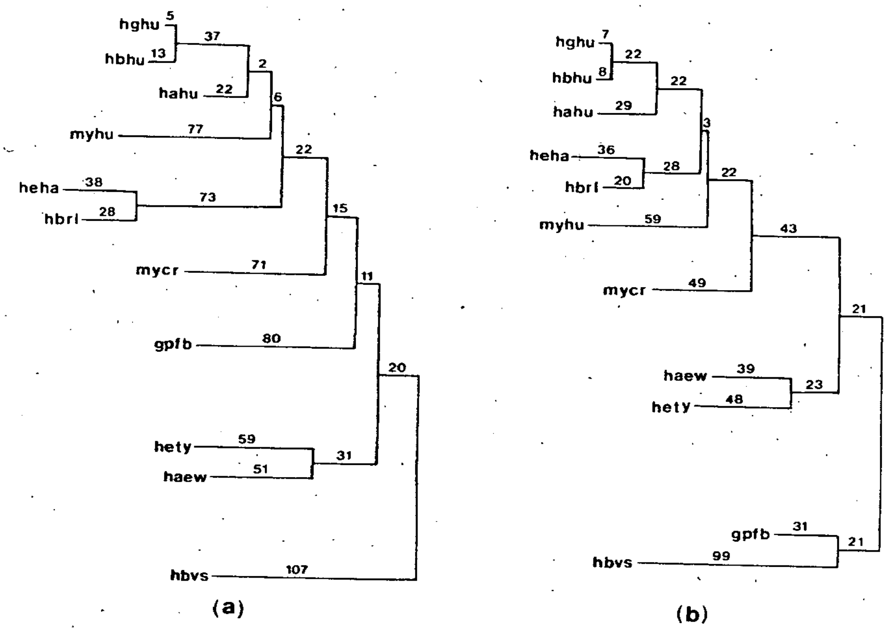


Fig. 4. Phylogenetic trees for 11 globin sequences as determined by **a** simple pairwise alignments and **b** the multiple alignments shown in Fig. 5. The four-letter designations are hghu, human globin γ chain; hbhu, human globin β chain; hahu, human globin α chain; heha, hagfish hemoglobin; hbrl, lamprey hemoglobin; myhu, human myoglobin; mycr, gastropod myoglobin; hety, earthworm hemoglobin (*Tylorrhynchus*); haew, earthworm hemoglobin (*Lumbricus*); gpfb, kidney bean leghemoglobin; hbvs, bacterial hemoglobin (*Vitreoscilla*). The same designations are used in Fig. 5 and Table 2.

to the globins of higher invertebrates and vertebrates than are those from annelid worms (Fig. 4a). Again, a more traditional grouping is obtained with the progressive alignment procedure (Fig. 4b). The multiple alignment generated by the procedure (Fig. 5) appears to be an accurate depiction of the history of events during globin evolution, and the degrees of similarity of the various globins based on these alignments are also more in line with expectations than are those found from simple binary alignments (Table 2).

Tyrosine Kinase-like Sequences

We had previously aligned a set of nine tyrosine kinase-like sequences and constructed a tree based on a simple pairwise matrix (Feng et al. 1985), and it was naturally of interest to see how the progressive alignment treatment compared (Table 3). In this case, unlike the situations with the superoxide dismutases and hemoglobins, the branching orders found by the two procedures did not differ (Fig. 6a and b). The multiple sequence alignment that was generated automatically during the procedure (Fig. 7) was somewhat different from the "eyeball" alignment made previously on the basis of a series of pairwise comparisons, although the same 14 invariant residues occur coincidentally in both renditions (Fig. 7). The trees themselves are not significantly different, although the branch lengths differ slightly.

Discussion

The concept of using pairwise alignments iteratively to establish phylogenetic relationships is hardly new. Moore et al. (1973) constructed the best possible dendrogram for a set of sequences by an iterative

```
                                                     *     *
hghu                GHFTEEDKATI  TSLW  GKV  NVEDAGGETLGRLLVVYPWTQRFFDSFGNLSSASAIMGNPK VKAHGKKVLTSLG
hbhu                VHLTPEEKSAV  TALW  GKV  NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK VKAHGKKVLGAFS
hahu                 VLSPADKTNV  KAAW  GKVGAHAGEYGAEALERMFLSFPTTKTYFPHF DLSH    GSAQ VKGHGKKVADALT
heha         PITDHGQPPTLSEGDKKAI  RESW  PQIYKNFEQNSLAVLLEFLKKFPKAQDSFPKFSAKKS  HLEQDPA VKLQAEVIINAVN
hbrl         PIVDSGSVAPLSAAEKTKI  RSAW  APVYSNYETSGVDILVKFFTSTPAAQEFFPKFKGMTSADQLKKSAD VRWHAERIINAVN
myhu                 GLSDGEWQLV  LNVW  GKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASED LKKHGATVLTALG
mycr                 SLQPASKSAL  ASSWKTLAKDAATIQNNGATLFSLLFKQFPDTRNYFTHFGNM SDAEMKTTGV GKAHSMAVFAGIG
haew                 KKQCGVLEGLKVKSEWGRAYGSGHDREAFSQAIWRATFAQVPESRSLFKR     VHGDHTSDPA FIAHAERVLGGLD
hety                  TDCGILQRILVLQQWAQVYSVGESRTDFAIDVFNNFFRTNPD RSLFNR     VNGDNVYSPE FKAHMVRVFAGFD
gpfb                  GAFTEKQEALVNSSW EAFK  GNIPQYSVVFYTSILEKAPAAKNLFSF    LANGVDPTNPK LTAHAESLFGLVR
hbvs                   MLDQQTINIIKATV PVLK  EHGVTITTTFYKNLFAKHPEVRPLFD     MGRQESLEQPKALAMTVLAAAQNIE

hghu         DAIKHLD   DLKGTFAQLSELHCDKLHVDPENFKLLGNVLVTVLAIHFGKEFTPEVQASWQKMV  TGVASALSSRYH
hbhu         DGLAHLD   NLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVV  AGVANALAHKYH
hahu         NAVAHVD   DMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFL  ASVSTVLTSKYR
heha         HTIGLMDKEAAMKKYLKDLSTKHSTEFQVNPDMFKELSAVFVSTM          GGKAAYEKLF  SIIATLLRSTYDA
hbrl         DAVASMDDTEKMSMKLRDLSGKHAKSFQVDPQYFKVLAAVIADTV          AAGDAGFEKLM  SMICILLRSAY
myhu         GILKKKGHHE  AEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKAL  ELFRKDMASNYKE LGFQG
mycr         SMIDSMDDADCMNGLALKLSRNHIQRKIGASRFGE MRQVFPNFLDEALGGGASGDVKGAWDALL  AYLQDNKQA  QA L
haew         IAISTLDQPATLKEELDHLQVQHEGRKIPDNYFDA FKTAILHVVAAQLGERCYSNNEEIHDAIACDGFARVLPQVLERG IKGHH
hety         ILISVLDDKPVLDQALAHYAAFH  LQFGTIPFKA FGQTMFQTIAEHI      HGADIGAWRAC  YA   EQIVT G ITA
gpfb         DSAAQLRANGAVVAD AALGSIHSQKGVSNDQFLV VKEALLKTLKQAV      GDKWTDQLSTALELA  YDELAAAI KKAYA
hbvs         NLPAILPAVKKIAVKHCQAGVAAAHYPIVGQELLGAIKEVLGDAATDDI      LDAWGKAYGVIADV  FIQVEADLYAQAVE
```

Fig. 5. Multiple alignment of 11 globins determined by progressive method. Asterisks denote locations where all 11 residues are identical. The order of the sequences was strictly based on the permutative trial described in the text. See legend to Fig. 4 for four-letter designations.

Table 2. Percent identities calculated from binary (upper triangle) and progressive (lower triangle) alignment methods

	Globins										
	hghu	hbhu	hahu	heha	hbrl	myhu	mycr	haew	hety	gpfb	hbvs
hghu		73	42	29	28	24	22	17	16	25	17
hbhu	73		45	26	24	25	22	18	18	23	20
hahu	42	45		20	35	27	24	22	19	15	18
heha	27	25	25		44	20	22	21	17	17	18
hbrl	26	24	34	44		23	19	26	23	15	14
myhu	25	25	28	18	23		22	21	17	17	19
mycr	21	21	23	19	18	21		18	20	15	25
haew	17	14	15	15	15	12	18		34	20	16
hety	16	15	14	12	14	12	18	34		16	13
gpfb	17	19	14	18	16	14	15	17	15		24
hbvs	11	11	12	6	10	10	10	11	12	15	

pairwise process, and, more recently, Hogeweg and Hesper (1984) used a heuristic approach for generating trees that also depends on successive pairwise alignments. As far as we know, however, the notion of "once a gap, always a gap," coupled with progressive pairwise alignment, has not been utilized before. Gap preservation is achieved by the insertion of neutral elements that hold the gap positions fixed during each progressive realignment.

Two things are certain: the method, while heuristic, provides multiple sequence alignments that are based on objective criteria, and trees derived from these alignments appear to be in harmony with the biology of the proteins as evidenced by the phylogeny of the organisms from which they are obtained. The simplicity of the procedure is attested to by the small number of pairwise comparisons that must be undertaken to produce the multiple alignment (Table 4). Thus, if 10 sequences are to be aligned, only 61 comparisons have to be made. This is a smaller number of alignments than is ordinarily performed when a set of jumbles is made for a single quantitative alignment. In this regard, we have eschewed the use of jumbled comparisons in the initial alignments in favor of an empirically determined average random score.

Kinds of Sequence Alignment

Broadly speaking, there are three kinds of multiple sequence alignment: (1) structural equivalence types, (2) global optimization methods, and (3) historical alignments. The first of these, structural equivalence, is used mainly by crystallographers. The goal is to align those segments of two protein sequences that occupy equivalent three-dimensional orientations. As such, these studies are usually restricted to protein families at least one member of which has had an x-ray structure determined (Bajaj and Blundell 1984). The interest is focused on present-

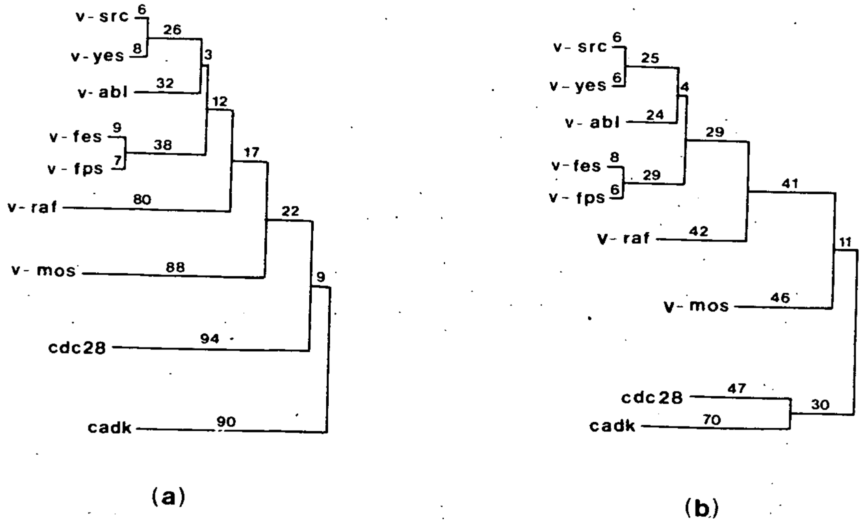


Fig. 6. Phylogenetic trees for nine tyrosine kinase-like sequences determined from a simple pairwise alignments and **b** progressive alignment. The four-letter designations are v-src, avian Rous sarcoma virus transforming factor; v-yes, avian Y73 sarcoma virus transforming factor; v-abl, Abelson murine leukemia virus transforming factor; v-fes, feline sarcoma virus transforming factor; v-fps, avian Fujinami virus transforming factor; v-raf, murine retroviral transforming factor; v-mos, mouse sarcoma virus transforming factor; cdc28, yeast cell division control factor; cadk, bovine cyclic AMP-dependent kinase. The same four-letter designations are used in Fig. 7 and Table 3.

```
                                                         *  *  *  *
v-src                           GLAK  DAW EIPRESLRLEAKLGQGCFGEVWMG   TWND T TRVAIKTLKPG TMSP EAFLQEA
v-yes                           GLAK  DAW EIPRESLRLEVKLGQGCFGEVWMG   TWNG T TKVAIKTLKLG TMMP EAFLQEA
v-abl                     TIYGVSPNYDKW EMERTDITMKHKLGGGQYGEVYEG   VWKKYS LTVAVKTLKED TMEV EEFLKEA
v-fes                       VLNRAVPKDKW VLNHEDLVLGEQIGRGNFGEVFSG   RLRADN TLVAVKSCRETLPPDIKAKFLQEA
v-fps                       VLTRAVLKDKW VLNHEDVLLGERIGRGNFGEVFSG   RLRADN TPVAVKSCRETLPPELKAKFLQEA
v-raf                          SSYY   W KMEASEVMLSTRIGSGSFGTVYKG   KWHGD  VAVKILKVVDPTPEQLQA FRNEV
v-mos                         GLPRRLAWFSIDWEQVCLMHRLGSGGFGSVYKA   TYHGVP VAIKQVNKCTEDLRASQRSFWAEL
cdc28                            MSGELANYKR      LEKVGEGTYGVVYKALDLRPGQGQRVVALKKIRLESEDEGVPSTAIREIS
cadk      LAKAKEDFLKKWENPAQNTAHLDQFER      IKTLGTGSFGRVMLVKHMETGNHYAMKILDKQKVVKLKQ  IEHTLNEKR

v-src     QV  MKKLRHEKLVQLYAV  VSEEPI YIVIEYMSKGSL   LDFLKGEM   GKYLRLPQLVDMAAQIASGMAYVERMNY
v-yes     QI  MKKLRHDKLVPLYAV  VSEEPI YIVTEFMTKGSL   LDFLKEGE   GKFLKLPQLVDMAAQIADGMAYIERMNY
v-abl     AV  MKEIKHPNLVQLLGVC TREPPF YIITEFMTYGNL   LDYLRECN   RQEVSAVVLLYMATQISSAMEYLEKKNF
v-fes     KI  LKQYSHPNIVRLIGVC TQKQPI YIVMELVQGGDF   LTFLRT E   GARLRMKTLLQMVGDAAAGMEYLESKCC
v-fps     RI  LKQCNHPNIVRLIGVC TQKQPI YIVMELVQGGDF   LSFLRS K   GPRLKMKKLIKMMENAAAGMEYLESKHC
v-raf     AV  LRKTRHVNILLFMG Y MTKDNL AIVTQWCEGSSL   YKHLHV Q   ETKFQMFQLIDIARQTAQGMDYLHAKNI
v-mos     NIAGLRHDNIVRVVAASTRTPEDSNSLGTIIMEFGGNVTLHQVIYDATRSPEPLSCRKQLSLGKCLKYSLDVVNGLLFLHSQSI
cdc28     LLKELKDDNIVRLYDI   VHSDAHKL YLVFEF LDLDLKRYM EGIPKDQPLG   ADIVK  KFMMQLCKGIAYCHSHRI
cadk      ILQAVNFPFLVKL EF   SFKDNSNL YMVMEY VP GGEMFS H LRRIGRFS    EPHAR  FYAAQIVLTFEYLHSLDL

             *    *               ***                              ***               *    *
v-src     VHRDLRAANILVGENLVCKVADFGLARLIEDNEYTARQGAK FPIK WTAPEAA   LYGRFTIKSDVWSFGILLTELTTKGRV
v-yes     IHRDLRAANILVGDNLVCKIADFGLARLIEDNEYTARQGAK FPIK WTAPEAA   LYGRFTIKSDVWSFGILLTELVTKGRV
v-abl     IHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTAHAGAK FPIK WTAPESL   AYNKFSIKSDVWAFGVLLWEIATYGMS
v-fes     IHRDLAARNCLVTEKNVLKISDFGMSREAADGIYAASGGLRQVPVK WTAPEAL   NYGRYSSESDVWSFGILLWETFSLGAS
v-fps     IHRDLAARNCLVTEKNTLKISDFGMSRQEEDGVYASTGGMKQIPVK WTAPEAL   NYGWYSSESDVWSFGILLWEAFSLGAV
v-raf     IHRDMKSNNIFLHEGLTVKIGDFGLATVKSRWSGSQQVEQPTGSVL WMAPEVIRMQDDNPFSFQSDVYSYGIVLYELMA GEL
v-mos     LHLDLKPANILISEQDVCKISDFGCSQKLQDLRGRQASPPHIGGTYTHQAPEILKGEIATP   KADIYSFGITLWQMTT REV
cdc28     LHRDLKPQNLLINKDGNLKLGDFGLARAF GVPLRAYTHEIVTLWY  RAPEVLLGGKQYS  TGVDTWSIGCIFAEMCN RKP
cadk      IYRDLKPENLLIDQQGYIQVTDFGFAKR   VKGRTWTL CGTPEY  LAPEIIL SKGYN  KAVDWWALGVLIYEMAAGYPP

v-src     PYPGMVN R  EVLDQVERG   YRM PCP PECPESLHDLMCQCWRKDPEERPTFKYL       QAQLLPACVLEVAE
v-yes     PYPGMVN R  EVLEQVERG   YRM PCP QGCPESLHELMKLCWKKDPDERPTFEYI       QSFLEDYFTAAEPSGY
v-abl     PYPGIDL S  QVYELLEKD   YRM ERP EGCPEKVYELMRACWQWNPSDRPSFAEI       HQAFETMFQESSIS
v-fes     PYPNLSN Q  QTREFVEKG   GRL PCP ELCPDAVFRLMEQCWAYEPGQRPSFSAI       YQELQSIRKRHR
v-fps     PYANLSN Q  QTREAIEQG   VRL EPP EQCPEDVYRLMQRCWEYDPHRRPSFGAV       HQDLIAIRKRHR
v-raf     PYAHINNRD  QIIFMVGRGYASPDL SRLYKNCPKAIKRLVADCVKKVKEERPLFPQILSSIELLQHSLPKINRSAPE
v-mos     PYS  GEPQ  YVQYAVVAYNLRPSLAGAVFTASLTG  KALQNIIQSCWEARGL   QRPSAELLHRDL KAFRGTLG
cdc28     IFSGDSEIDQIFKIFRVLGTPNEAIWPDIVYLPDFKPSFPQWRR KDLSQVVPSL   DPRGIDLLDKLLAYDPINRISA
cadk      FFADQP I QIYE  KIVSG       KVRFPSHFSSDLKDLLR  NLLQV  DL   TKRFGNLKDGVNDIKNHKWF
```

Fig. 7. Multiple alignment of seven tyrosine kinase-like oncogene sequences and those of yeast cdc28 and bovine heart cyclic AMP-dependent kinase as determined by progressive method. Asterisks denote locations where all nine residues are identical.

day structure without regard for how the structures came to be.

Global optimization methods are designed to accommodate a set of sequences in a multiple alignment that maximizes overall similarity. Three-dimensional extensions of the Needleman-Wunsch algorithm, for example, have been used to achieve such alignments (Jue et al. 1980; Murata et al. 1985), and Johnson and Doolittle (1986) have used the overlapping approach pioneered by Fitch (1966, 1970) to generate four-way and five-way alignments. Again, these alignments are made without regard to historical detail.

Historical alignments are based on the notion that divergent evolution is fundamentally binary in nature. Long ago Dayhoff et al. (1972), noting that matrix methods greatly foreshorten the more ancient branches in evolutionary trees, used a common-

Table 3. Percent identities calculated from binary (upper triangle) and progressive (lower triangle) alignment methods

	Tyrosine kinase-like sequences								
	v-src	v-yes	v-abl	v-fes	v-fps	v-raf	v-mos	cdc28	cadk
v-src		84	47	43	41	32	30	25	26
v-yes	84		49	43	41	35	30	25	26
v-abl	47	49		41	41	28	24	25	24
v-fes	41	43	41		79	30	26	25	26
v-fps	40	41	41	79		30	29	27	27
v-raf	30	31	26	29	30		27	23	23
v-mos	23	24	20	22	24	26		25	19
cdc28	18	19	16	16	17	21	25		26
cadk	19	18	17	15	16	18	20	26	

Table 4. Numbers of pairwise alignments required to construct a phylogenetic tree by a progressive method[a]

Number of sequences	Initial pairwise alignments	Additional iterative alignments	Total
3	3	2	5
4	6	4	10
5	10	6	16
6	15	8	23
7	21	10	31
8	28	12	40
9	36	14	50
10	45	16	61
11	55	18	73

[a] Values are minimal numbers for simple trees; compound trees need an additional alignment for each subcluster. Also, occasional negative segments in some trees will necessitate additional alignments

ancestor approach to alignment and tree building that was historical in principle. The character-based approach that they used was much clumsier than matrix methods, however, and eventually was abandoned. Subsequently, Holmquist (1979, p 939) drew attention to the fact that parsimony methods err significantly, "the magnitude of the error increasing with the distance of the nodal sequence from the present," and, more recently, Penny and Hendy (1986) have expounded on the theme that the minimal tree cannot be the historical tree.

It is obvious that methods based on mere global optimization will consistently underestimate evolutionary distances among the least related members of the set, striving as they do to achieve maximum alignment scores. The need is to throttle the tendency for optimization while preserving the notion of similar residues replacing one another. The progressive alignment procedure presented here appears to achieve that end. In its favor, the trees generated from these alignments appear to be in accord with biological expectations.

Superoxide Dismutase Relationships

The copper–zinc superoxide dismutase sequences have been the subject of much debate since the possibility was raised that the sequence found in the prokaryote *Photobacterium leiognathi* might be the result of a horizontal gene transfer from its ponyfish host (Martin and Fridovich 1981). Although solid evidence to the contrary was provided by Steffens et al. (1983), the notion has refused to go away (Bannister and Parker 1985). Our thinking about this matter is wholly in accord with that recently expressed by Leunissen and De Jong (1986): to wit, there is no basis for supposing anything other than a conventional history of events. Indeed, either of the evolutionary trees in Fig. 2 ought to dispel thoughts of a horizontal gene transfer for this gene, the photobacter position being entirely consistent with what would be expected for a typical prokaryotic–eukaryotic divergence. On the other hand, the tree made from pairwise alignments (2a) does have an unreasonble arrangement for the fruitfly and yeast, whereas the progressive tree is quite in line with conventional phylogeny.

It should be pointed out in passing that an apparent speed-up in the rate of copper–zinc superoxide dismutase evolution has occurred among the vertebrates (Lee et al. 1985). Thus, the apparent differences between mammalian and *Drosophila* sequences are much greater than would be expected on the basis of a comparison of the *Drosophila* and yeast sequences. The fact that there appears to have been a relaxation of selection pressures on the vertebrate superoxide dismutase should not affect the branching order, of course.

Hemoglobins and Myoglobins

The progressive alignment scheme also yields reasonable results when applied to distantly related globin sequences. In contrast to phylogenies employing a maximum parsimony method (Goodman et al. 1974), the progressive method roots the lamprey

and hagfish globins to the same branch as other vertebrate hemoglobins. Interestingly, an early study employing the common ancestor approach (Dayhoff and Eck 1968) also had the lamprey in this position. With regard to the relationship of animal and plant globins, the depth of the differences warrants a good deal of caution. Nonetheless, the recently published bacterial globin sequence (Wakabayashi et al. 1986) resembles the plant globins more than it does the animal globins, and it is not impossible that an unusual genetic event involving plants and symbiotic bacteria has occurred. A larger study encompassing all the known invertebrate and plant globin sequences may reveal more about the evolutionary connections of these proteins.

Concluding Remark

It is not our intention to reopen past skirmishing about the relative merits of strict parsimony methods and alternative treatments (Fitch 1981; Holmquist and Jukes 1981). Nor is it our aim merely to add one more comment to the enormous literature on the construction of evolutionary trees with sequence data (Tateno et al. 1982; Hogeweg and Hesper 1984; Penny and Hendy 1986, to name but a few). Rather, we simply offer a heuristic procedure for a computer-determined multiple alignment of related amino acid sequences that can be effected rapidly by objective criteria. Evolutionary trees drawn directly from these alignments appear to be very much in accord with biological expectations.

Acknowledgments. We acknowledge many helpful discussions with Mark Johnson and Marcella McClure; we are especially grateful to Mark Johnson for writing the programs BORD and TREEplot. This work was supported by NIH Grant GM-34434 and a grant from the American Cancer Society.

Note Added in Proof. During the period since the acceptance of this article we have applied the procedure in numerous settings, and, in some cases, the final alignment was slightly imperfect. The situation was remedied, however, by aligning each new sequence, or set of sequences, with an average sequence of all the sequences already aligned. This was accomplished by simply looking up the matrix value for every pair of residues at each position and averaging them. We are grateful to Steve Hanks for bringing the problem to our attention and to Mark Johnson for helping with the solution.

References

Bajaj M, Blundell T (1984) Evolution and the tertiary structure of proteins. Ann Rev Biophys Bioeng 13:453–492

Bannister JV, Parker MW (1985) The presence of a copper/zinc superoxide dismutase in the bacterium *Photobacterium leiognathi*: a likely case of gene transfer from eukaryotes to prokaryotes. Proc Natl Acad Sci USA 82:149–152

Cannon RE, White JA, Scandalios JG (1987) Cloning of cDNA for maize superoxide dismutase 2 (SOD2). Proc Natl Acad Sci USA 84:179–183

Dayhoff MO, Eck RV (1968) Atlas of protein sequence and structure 1967–1968. National Biomedical Research Foundation, Silver Spring MD, p 19

Dayhoff MO, Park CM, McLaughlin PJ (1972) Building a phylogenetic tree: cytochrome c. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5. National Biomedical Research Foundation, Washington DC, pp 7–16

Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model for evolutionary change. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 3. National Biomedical Research Foundation, Washington DC, pp 345–358

Doolittle RF (1981) Similar amino acid sequences: chance or common ancestry? Science 214:149–159

Feng DF, Johnson MS, Doolittle RF (1985) Aligning amino acid sequences: comparison of commonly used methods. J Mol Evol 21:112–125

Fitch WM (1966) An improved method of testing for evolutionary homology. J Mol Biol 16:9–16

Fitch WM (1970) Further improvements in the method of testing for evolutionary homology among proteins. J Mol Biol 49:1–14

Fitch WM (1977) On the problem of discovering the most parsimonious tree. Am Nat 111:223–257

Fitch WM (1981) The old REH theory remains unsatisfactory and the new REH theory is probiematical—a reply to Holmquist and Jukes. J Mol Evol 18:60–67

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 15:279–284

Fredman ML (1984) Computing evolutionary similarity measures with length independent gap penalties. Bull Math Biol 46:553–566

Goodman M, Moore GW, Barnabas J, Matsuda G (1974) The phylogeny of human globin genes investigated by the maximum parsimony method. J Mol Evol 3:1–48

Hogeweg P, Hesper B (1984) The alignment of sets of sequences and the construction of phyletic trees: an integrated method. J Mol Evol 20:175–186

Holmquist R (1979) The method of parsimony: an experimental test and theoretical analysis of the adequacy of molecular restoration studies. J Mol Biol 135:939–958

Holmquist R, Jukes T (1981) The current status of REH theory. Reply to an essay by Fitch. J Mol Evol 18:47–59

Hunt LT, Hurst-Calderone S, Dayhoff MO (1978) Globins. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 3. National Biomedical Research Foundation, Washington DC, pp 229–249

Jabusch JR, Farb DL, Kerschensteiner DA, Deutsch HF (1980) Some sulfhydryl properties and primary structure of human superoxide dismutase. Biochemistry 19:2310–2316

Johansen JT, Overballe-Petersen C, Martin B, Hasemann B, Svendsen I (1979) The complete amino acid sequence of copper–zinc superoxide dismutase from *Saccharomyces cerevisiae*. Carlsberg Res Commun 44:201–217

Johnson MS, Doolittle RF (1986) A method for the simultaneous alignment of three or more amino acid sequences. J Mol Evol 23:267–278

Jue RA, Woodbury NW, Doolittle RF (1980) Sequence homologies among *E. coli* ribosomal proteins: evidence for evolutionarily related groupings and internal duplications. J Mol Evol 15:129–148

Kernighan BW, Ritchie DM (1978) The C programming language. Prentice-Hall, Englewood Cliffs NJ

Klotz LC, Blanken RL (1981) A practical method for calculating evolutionary trees from sequence data. J Theor Biol 91:261–272

Lee YM, Friedman DJ, Ayala FJ (1985) Superoxide dismutase: an evolutionary puzzle. Proc Natl Acad Sci USA 82:824–828

Leunissen JAM, De Jong WW (1986) Copper/zinc superoxide dismutase: how likely is gene transfer from ponyfish to *Photobacterium leiognathi*? J Mol Evol 23:250–258

Martin JP, Fridovich I (1981) Evidence for a natural gene transfer from the ponyfish to its bioluminescent bacterial symbiont *Photobacter leiognathi*. J Biol Chem 256:6080–6089

Moore GM, Goodman M, Barnabas J (1973) An iterative approach from the standpoint of the additive hypothesis to the dendrogram problem posed by molecular data sets. J Theor Biol 38:423–457

Murata M, Richardson JS, Sussman JL (1985) Simultaneous comparison of three protein sequences. Proc Natl Acad Sci USA 82:3073–3077

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Penny D, Hendy M (1986) Estimating the reliability of evolutionary trees. Mol Biol Evol 3:403–417

Rocha HA, Bannister WH, Bannister JV (1984) The amino-acid sequence of copper/zinc superoxide dismutase from swordfish liver. Eur J Biochem 145:477–484

Sankoff D, Cedergren RJ, McKay WM (1982) A strategy for sequence phylogeny research. Nucleic Acids Res 10:421–431

Sellers PH (1974) Evolutionary distances. SIAM J Appl Math 26:787–793

Steffens GJ, Bannister JV, Bannister WH, Flohe L, Gunzler WA, Kim S-MA, Otting F (1983) The primary structure of Cu-Zn superoxide dismutase from *Photobacterium leiognathi*: evidence for a separate evolution of Cu–Zn superoxide dismutase in bacteria. Hoppe-Seyler's Z Physiol Chem 364:675–690

Steinman HM, Naik VR, Abernathy JL, Hill RL (1974) Bovine erythrocyte superoxide dismutase J Biol Chem 249:7326–7338

Tateno Y, Nei M, Tajima F (1982) Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. J Mol Evol 18:387–404

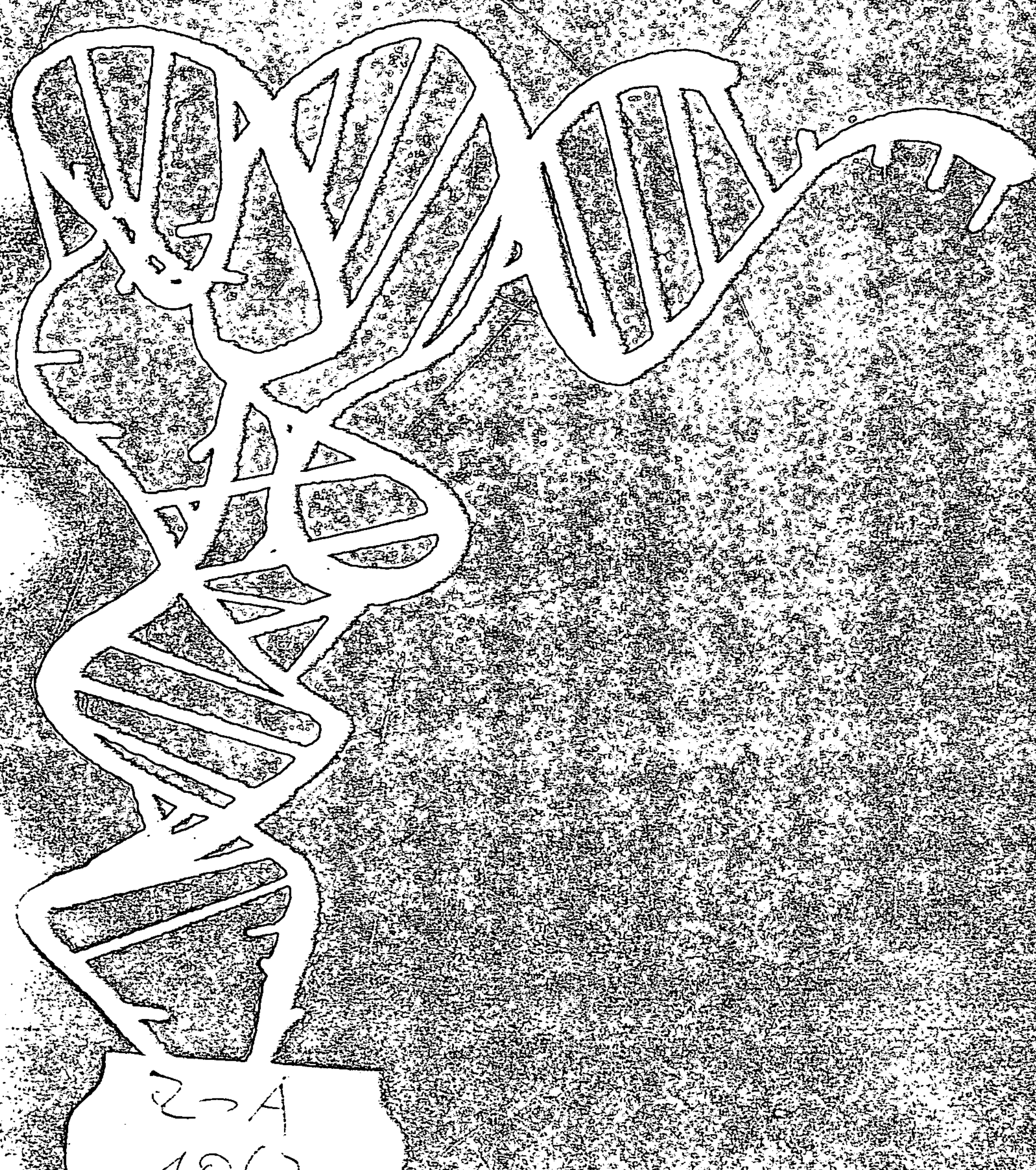
Wakabayashi S, Matsubara H, Webster DA (1986) Primary sequence of a dimeric bacterial hemoglobin from *Vitreoscilla*. Nature 322:481–483

Zelenik M, Rudloff V, Braunitzer G (1979) Die Aminosauresequenz des monmeren Hamoglobins von Lampetra fluviatilis. Hoppe-Seyler's Z Physiol Chem 360:1879–1894

Received December 11, 1986/Revised January 21, 1987

cleic Acids Research

Volume 12 number 1 Part 1 January 11 1984



IRL PRESS

ISSN 0305 1048

Volume 12 Number 1 1984 Nucleic Acids Research

A comprehensive set of sequence analysis programs for the VAX

John Devereux, Paul Haeberli* and Oliver Smithies

Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, USA

Received 18 August 1983

ABSTRACT

The University of Wisconsin Genetics Computer Group (UWGCG) has been organized to develop computational tools for the analysis and publication of biological sequence data. A group of programs that will interact with each other has been developed for the Digital Equipment Corporation VAX computer using the VMS operating system. The programs available and the conditions for transfer are described.

INTRODUCTION

The rapid advances in the field of molecular genetics and DNA sequencing have made it imperative for many laboratories to use computers to analyze and manage sequence data. UWGCG was founded when it became clear to several faculty members at the University of Wisconsin that the there was no set of sequence analysis programs that could be used together as a coherent system and be modified easily in response to new ideas.

With intramural support a computer group was organized to build a strong foundation of software upon which future programs in molecular genetics could be based. This initial project has been completed and the resulting programs, written in Fortran 77, are available for VAX computers using the VMS operating system. Most of the programs can be used with only a terminal, although several require a Hewlett Packard plotter.

UWGCG software has been installed for testing at eight different institutions. A simple method has been developed for transferring and maintaining this system on other VAX computers.

DESIGN PRINCIPLES

UWGCG program design is based on the "software tools" approach of Kernighan and Plauger(1). Each program performs a simple function and is easy to use. The programs can be used independently in different combinations so

that complex problems are solved by the use of several programs in succession. New programming is simplified since less effort is required to bridge a gap between existing programs.

UWGCG software is designed to be maintained and modified at sites other than the University of Wisconsin. The program manual is extensive and the source codes are organized to make modification convenient. Scientists using UWGCG software are encouraged to use existing programs as a framework for developing new ones. Our copyright can be removed from any program modified by more than 25% of our original effort.

PROGRAMS AVAILABLE FROM UWGCG

The programs described below are named and defined individually in Table 1. Program names in the text are underlined.

Comparisons

Comparisons may be done with "dot plots" using the method of Maizel and Lenk(2). Optimal alignments can be generated by the methods of Needleman and Wunsch(3), of Sellers(4), and the "local homology" method of Smith and Waterman(5). The Smith and Waterman alignment algorithm is also the most sensitive method available for identifying similarities between weakly related sequences.

Mapping and Searching

Mapping is available in several formats. Graphic maps display all of the cuts for each restriction enzyme on parallel lines. This graphic map facilitates selection of enzymes for isolating any region of a sequenced DNA molecule. Sorted maps in tabular format arrange the fragments from any digestion in order of molecular weight to show which fragments are similar in size and thus likely to be confused in gels. Another frequently used mapping format, designed by Frederick Blattner(6), displays the enzyme cuts above the original DNA sequence. Both strands of the DNA and all six frames of translation are shown.

All mapping programs will search for user-specified sequences, allowing features to be marked at the appropriate position on a restriction map. The mapping and searching programs can be used to aid site-specific mutagenesis experiments by showing where mutations could generate new restriction sites. All of the positions in a sequence where a synthetic probe could pair with one or more mismatches can also be located. Sequences related to less precisely defined features such as promoters or intervening sequence splice sites, can be located with a program that uses a consensus sequence as a probe. The

Table 1

Programs Available from UWGCG

Name	Function
DotPlot+	makes a dot plot by method of Maizel and Lenk(2)
Gap	finds optimal alignment by method of Needleman and Wunsch(3)
BestFit	finds optimal alignment by method of Smith and Waterman(5)
MapPlot+	shows restriction map for each enzyme graphically
MapSort	tabulates maps sorted by fragment position and size
Map	displays restriction sites and protein translations above and below the original sequence(Blattner,6)
Consensus	creates a consensus table from pre-aligned sequences
FitConsensus	finds sequences similar to a consensus sequence using a consensus table as a probe
Find	finds sites specified interactively
Stemloop	finds all possible stems (inverted repeats) and loops
Fold*	finds an RNA secondary structure of minimum free energy by the method of Zuker(7)
CodonPreference+	plots the similarity between the codon choices in each reading frame and a codon frequency table(8)
CodonFrequency	tabulates codon frequencies
Correspond	finds similar patterns of codon choice by comparing codon frequency tables (Grantham et al,9)
TestCode+	finds possible coding regions by plotting the "TestCode" statistic of Fickett(10)
Frame+	plots rare codons and open reading frames(8)
PlotStatistics+	plots asymmetries of composition for one strand
Composition	measures composition, di and trinucleotide frequencies
Repeat	finds repeats (direct, not inverted)
Fingerprint	shows the labelled fragments expected for an RNA fingerprint
Seqed	screen oriented sequence editor for entering, editing and checking sequences
Assemble	joins sequences together
Shuffle	randomizes a sequence maintaining composition
Reverse	reverses and/or complements a sequence
Reformat	converts a sequence file from one format to another
Translate	translates a nucleotide into a peptide sequence
BackTranslate	translates a peptide into a nucleotide sequence
Spew	sends a sequence to another computer
GetSeq	accepts a sequence from another computer
Crypt	encrypts a file for access only by password
Simplify	substitutes one of six chemically similar amino acid families for each residue in a peptide sequence
Publish	arranges sequences for publication
Poster+	plots text (for labelling figures and posters)
OverPrint	prints darkened text for figures with a daisy wheel printer

+ requires a Hewlett Packard Series 7221 terminal plotter
* Fold is distributed by Dr. Michael Zuker not UWGCG.

mapping programs can also be used on protein sequences to identify the peptides resulting from proteolytic cleavage.

Secondary Structure

Three programs are available to examine secondary structure in nucleic acids. The program StemLoop identifies all inverted repeats. An implementation of Dr. Michael Zuker's Fold program(7) finds an RNA secondary structure of minimum free energy based on published values of stacking and loop destabilizing energies. The "dot plot" comparison (mentioned above) of a sequence compared to its opposite strand gives a graphic picture of the pattern of inverted repeats in a sequence.

Analysis of Composition and the Location of Genetic Domains

Regions of a sequence with non-random base distribution can be displayed with three graphic tools designed to identify genetic domains. The program CodonPreference(8) identifies potential coding regions by searching through each reading frame for a pattern of preferred codon choices. The CodonPreference plot predicts the level of translational expression of mRNAs and helps identify frame shifts in DNA sequence data. Patterns of codon choice can be compared with the program Correspond(9). When a strong pattern of codon preferences is not expected, the "TestCode" statistic of Fickett(10) can be plotted to show regions of compositional constraint at every third base. Another program plots asymmetries of composition by strand. Strand asymmetries have been associated with genetic domains by several authors(11)(12). A fourth program called Frame marks the positions of rare codons and open reading frames on a graph showing all six reading frames.

Several tools are available to measure content and to count dinucleotide, trinucleotide, neighbor and repeat frequencies. A program that predicts RNA fingerprint patterns and another that tabulates codon frequencies complete the group of programs that analyze composition.

Sequence Manipulation

Sequences may be entered, assembled, edited, reversed, randomized, reformatted, translated, back-translated, documented, transferred, or encrypted rapidly with a large set of sequence manipulation tools.

A screen-oriented editor is available that allows sequences to be entered and checked. After a sequence is entered, it may be reentered for proofreading. Whenever a reentered base is at variance with the original, the terminal bell rings and the position is marked. Existing sequences can be edited quickly by moving directly to a sequence position specified by either a coordinate or a sequence pattern. The program can reassign the terminal's

keys to place G, A, T and C conveniently under the fingers of one hand in the same order as the lanes of a sequencing gel.

Programs are available for changing sequence file format. Sequence data from any source can be used in UWGCG programs, and sequence files maintained with UWGCG software can be converted for use in other non-UWGCG programs. For instance, the programs of Roger Staden(13) or Intelligenetics Inc.(14) could be used to assemble a sequence from the sequences of many small sub-fragments generated by DNAase I digestion. The assembled sequence could then be reformatted for use in any UWGCG program. A program is available that transfers sequences to and from other computers.

Sequence Publication

A program, Publish, will format sequences into figures. Publish has alternatives for line size, numbering, scaling, translation and comparison to other sequences. Poster is a program that will plot text on figures.

GENERAL FEATURES OF UWGCG SOFTWARE

Interactive Style

Each program is run by simply typing its name. Every parameter required by the program is obtained interactively. Questions are answered with a file name, a yes, a no, a number, or a letter from a menu. Default answers are displayed. Programs are insensitive to absurd answers and will ask the question again if, for instance, you name a file that does not exist or if you use a nonnumeric character when typing a number. Special features such as plotting features oriented to publication, are obtained by using an extra word next to the program's name when the program is run. Thus parameter queries are kept to a minimum for the normal use of each program.

Data

Both the NIH-GenBank(15) and the EMBL(16) nucleotide sequence data libraries are available "on-line" to any UWGCG program. A Search utility will locate sequences in the libraries by key word. A Find utility will locate library entries containing any specified sequence. A program is available that installs the new data sent periodically from GenBank and EMBL to update their data libraries.

All of the data in the system are stored in text files that can be read and modified easily. Every data file has an English heading describing the contents. The data files may be copied by each user for analysis or modification. Programs recognize and read user-modified input data automatically. Data files can be modified with any text editor.

Sequence File Structure

Sequences are maintained in files that allow documentation and numbering both above and within the sequence. This file format is compatible with both of the nucleic acid sequence libraries and has been adopted as the standard sequence file format by the data base project at the European Molecular Biology Lab. Because genetic manipulations commonly involve linking several molecules of known sequence, UWGCG sequence files are designed to support concatenation by allowing comments to appear within the sequences at any location. Coding sequences or the boundaries between cloning vector and insert, for instance, can be marked within the sequence itself for immediate identification.

Sequence Symbols

All possible nucleotide ambiguities and all standard one-letter amino acid codes are part of the UWGCG symbol set that includes all alphabetic characters plus five additional characters. The proposed IUB-IUPAC standard nucleotide ambiguity symbols(17) are used for the mapping, searching and comparison programs. Lower case characters are used in sequences to indicate uncertainty as distinct from ambiguity. This allows the entire lexicon of symbols to be reused with same meaning, but with the prefix "maybe-." This reuse of the symbol set in lower case makes the uncertainty symbols more complete, understandable and visible.

Symbol Comparison

Sequence analysis programs generally make comparisons between sequence symbols (bases or amino acids) in order to find enzyme sites, create alignments, locate inverted repeats etc. These symbol comparisons are handled in several ways.

Symbol comparisons for alignment, comparison and secondary structure analysis are made by looking up a value in a symbol comparison table for the quality of the match. The table might contain 1's for matches and 0's for mismatches. If amino acids are being compared, however, a real number could be assigned at each position based on some previously assigned chemical similarity of the pair of residues or on the mutational distance between their codons. Standard symbol tables are provided by UWGCG, but the system is designed to allow each user to specify his own values.

Symbols comparisons for mapping and searching operations in nucleic acids are made by converting the IUB-IUPAC symbols into a binary code. The bits of this code represent G, A, T and C with ambiguity symbols causing more than one

bit to be set. A group of library functions identify overlap between the bits for each IUB-IUPAC symbol.

Documentation

Documentation is available both in printed form and on the terminal screen. A 350 page manual describes the operation of each program in detail, gives practical considerations and shows what will appear on the screen during a session with the program. Output files and plots are shown for the session. The data for the session shown in the documentation are included with the system so that the each program's operation can be checked. The "on-line" documentation is the same as the manual, but can be changed immediately when a program is modified.

All programs write output to files that are completely documented and sensibly organized for input to other programs. The input data, the program and the parameters used are clearly identified in every output file.

Procedure Library

UWGCG programs are written largely as calls to a library of 250 procedures designed to manipulate biological sequences. These procedures use data and file structures which have been designed to simplify program modification. For instance, standard operations such as reading sequences from files are always handled by a single library procedure. Thus a change in sequence file format requires only one subroutine to be modified for the new format to be acceptable to all of the programs in the system. Command procedures are available to help modify the library. The procedure library can be used by programs written in any language.

DISTRIBUTION OF UWGCG SOFTWARE

Intent

The intent of UWGCG is to make its software available at the lowest possible cost to as many scientists as possible.

Fees

A fee of $2,000 for non-profit institutions or $4,000 for industries is being charged for a tape and documentation for each computer on which UWGCG software is installed. While no continuing fee is required, UWGCG software, like the field it supports, is changing very rapidly. A consortium of industries and academic laboratories is planned to support the project in the future. The consortium will entitle its members to periodic updates and to influence the direction of new programming undertaken by UWGCG in return for a pledge of continuing financial support.

Copyrights

UWGCG retains the copyrights to all of its software and UWGCG must be contacted before all or any part of the its software package is copied or transferred to any machine. UWGCG is, however, mandated to provide research tools to help scientists working in the area of molecular genetics and we are glad to see our source codes become the basis of further programming efforts by other scientists. Copyright can be removed for any program modified by more than 25% of its original effort.

Tape Format

The UWGCG package is usually distributed in VAX/VMS "backup" format on a 9 track magnetic tape recorded at 1600 bits/inch. The system consists of about 1000 files using about 20,000 blocks at 512 bytes/block. The current versions of the GenBank and EMBL nucleotide sequence data bases are normally included which add another 3,000 files and require another 20,000 blocks.

Upon request UWGCG will make a card image tape of all of the Fortran 77 programs and procedures for reading on computers other than the VAX. The card image tape is usually provided at 1600 bits/inch with 80 characters/record and 10 records/block. Adaptation of UWGCG software to systems other than VAX/VMS may take considerable effort.

Equipment Required

UWGCG programs and command procedures will run on a Digital Equipment Corporation (DEC) VAX computer that is using version 3.0 or greater of the DEC VMS operating system. A tape drive is necessary; a floating point accelerator and a DEC Fortran compiler are helpful, but not required. All programs can be run from a DEC VT52 or VT100 terminal. Seven programs, as noted in table 1, require a Hewlett Packard 7221 terminal plotter wired in series with the terminal. Several utilities support a daisy wheel compatible printer attached to the terminal's pass-through port, however, all programs write output files suitable for printing on any standard device.

Inquiries

Inquiries may be sent to John Devereux at the Laboratory of Genetics, University of Wisconsin, Madison, WI, USA 53706, (608) 263-8970. UWGCG is not licensed to distribute Fold(7), but the UWGCG implementation is available from Michael Zuker, Division of Biological Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Canada, K1A OR6 (613) 992-4182.

ACKNOWLEDGEMENTS

UWGCG was started with software written for Oliver Smithies' laboratory

with NIH support from grants GM 20069 and AM 20120. UWGCG is directed by John Devereux and is operated as a part of the Laboratory of Genetics with the advice of a steering committee consisting of Richard Burgess, James Dahlberg, Walter Fitch, Oliver Smithies and Millard Susman. UWGCG is currently supported with intramural funds and with fees paid by the faculty and industries using the facility in Madison. This article is paper number 2684 from the Laboratory of Genetics, University of Wisconsin.

*Current address: Silicon Graphics Inc., 630 Clyde Court, Mountain View, CA 94043, USA

REFERENCES

1. Kernighan, B.W. and Plauger, P.J. (1976) Software Tools, Addison-Wesley Publishing Company, Reading, Massachusetts.
2. Maizel, J.V. and Lenk, R.P. (1981) Proceedings of the National Academy of Sciences USA 78, 7665-7669.
3. Needleman, S.B. and Wunsch, C.D. (1970) Journal of Molecular Biology 48, 443-453.
4. Sellers, P.H. (1974) SIAM Journal on Applied Mathematics 26, 787-793.
5. Smith, T.F. and Waterman, M.S. (1981) Advances in Applied Mathematics 2, 482-489.
6. Schroeder, J.L. and Blattner, F.R. (1982) Nucleic Acids Research 10, 69-84, Figure 1.
7. Zuker, M. and Stiegler, P. (1981) Nucleic Acids Research 9, 133-148.
8. Gribskov, M., Devereux, J. and Burgess, R.R. "The Codon Preference Plot: Graphic Analysis of Protein Coding Sequences and Gene Expression," submitted to Nucleic Acids Research.
9. Grantham, R. Gautier, C. Guoy, M. Jacobzone, M. and Mercier R. (1981) Nucleic Acids Research 9(1), r43-r74.
10. Fickett, J.W. (1982) Nucleic Acids Research 10, 5303-5318
11. Smithies, O., Engels, W.R., Devereux, J.R., Slightom, J.L., and S. Shen, (1981) Cell 26, 345-353.
12. Smith, T.F., Waterman, M.S. and Sadler, J.R. (1983) Nucleic Acids Research 11, 2205-2220.
13. Staden, R. (1980) Nucleic Acids Research 8, 3673-3694.
14. Clayton, J. and Kedes, L. (1982) Nucleic Acids Research 10, 305-321.
15. The GenBank(TM) Genetic Sequence Data Bank is available from Wayne Rindone, Bolt Beranek and Newman Inc., 10 Moulton Street, Cambridge, Massachusetts 02238, USA.
16. The EMBL Nucleotide Sequence Data Library is available from Greg Hamm, European Molecular Biology Laboratory, Postfach 10.2209, Meyerhofstrasse 1, 6900 Heidelberg, West Germany.
17. Personal communication from Dr. Richard Lathe, Transgene SA, 11 Rue Humann, 67000 Strasbourg, France.

Volume 15 Number 21 1987 Nucleic Acids Research

A comparison of eukaryotic viral 5′-leader sequences as enhancers of mRNA expression *in vivo*

Daniel R.Gallie, David E.Sleat, John W.Watts[1], Philip C.Turner[2] and T.Michael A.Wilson*

Departments of Virus Research and [1]Cell Biology, John Innes Institute, AFRC Institute of Plant Science Research, Colney Lane, Norwich NR4 7UH and [2]Department of Biochemistry, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

Received August 5, 1987; Accepted September 25, 1987

ABSTRACT

The 5'-untranslated leader sequences of several plant RNA viruses, and a portion of the 5′-leader of an animal retrovirus, were tested for their ability to enhance expression of contiguous open reading frames for chloramphenicol acetyltransferase (CAT) or β-glucuronidase (GUS) in tobacco mesophyll protoplasts, Escherichia coli and oocytes of Xenopus laevis. Translation of capped or uncapped transcripts was substantially enhanced in almost all systems by the leader sequence of either the U1 or SPS strain of TMV. All leader sequences, except that of TYMV, stimulated expression of 5′-capped GUS mRNA with the native prokaryotic initiation codon context, in electroporated protoplasts. Only the TMV leaders enhanced translation of uncapped GUS mRNAs in protoplasts and increased expression of uncapped CAT mRNA in microinjected X. laevis oocytes. In oocytes, the TYMV leader sequence was inhibitory.

In transformed E. coli, the TMV-U1 leader enhanced expression of both the native and eukaryotic context forms of GUS mRNA about 7.5-fold, despite the absence of a Shine-Dalgarno region in any of the transcripts. The absolute levels of GUS activity were all about 6-fold higher with mRNAs containing the native initiation codon context. In E. coli, the leaders of AlMV RNA4 and TYMV were moderately stimulatory whereas those of BMV RNA3, RSV and the SPS strain of TMV enhanced GUS expression by only 2- to 3-fold.

INTRODUCTION

Cis-acting features which influence the selection and translation of eukaryotic mRNAs are poorly understood. Surveys of sequences upstream from the AUG start codon have failed to identify a universal consensus sequence which might act as the eukaryotic equivalent of the prokaryotic Shine-Dalgarno region, a region essential for the expression of prokaryotic mRNAs in E. coli (1). Secondary structures within the 5′-untranslated leaders of some eukaryotic mRNAs have been claimed to promote

(2,3) or inhibit (4) translation initiation. In prokaryotic mRNAs, selection of start codons may also be influenced, in part, by low surrounding secondary structure (5). In the relaxed scanning model (6), 40S ribosomal subunits bind at the 5′-end of an mRNA and scan until the first AUG in the optimal context (5′-ACCAUGG-3′) is reached, at which point translation begins. Beyond this, little is known about the longer-range effects of specific sequences on expression of eukaryotic mRNA.

We have shown that translation of prokaryotic (7) and eukaryotic (8) mRNAs is greatly enhanced by a contiguous derivative of the 68-nucleotide, 5′-leader sequence of tobacco mosaic virus (TMV), U1 strain (called Omega (Ω); 9,10). The stimulatory effect of this Ω-like sequence (referred to as Ω′-U1) has been observed in vitro and in vivo, in both eukaryotic and prokaryotic translation systems. Tyc and co-workers (11) identified a second 80S ribosome binding site, centred on residues 14-16 (AUU) within Ω-U1 (or Ω′), which was upstream of, and in frame with, the predicted ribosome binding site at the first AUG codon (residues 68-70 in Ω-U1). The latter initiates synthesis of the 126,000 dalton (126Kd) protein encoded by TMV RNA. In Ω-U1 (and Ω′-U1), 51 nucleotides separate the AUU and AUG sequences which, in the presence of an inhibitor of elongation (sparsomycin), permit two ribosomes to bind simultaneously without steric hindrance. Initiation of translation of genomic TMV RNA under these conditions has been claimed to result in two unique dipeptides, Met-Thr and Met-Ala, (12) which may arise by illegitimate or legitimate initiation at the AUU and AUG sites, respectively. Yokoe and coworkers (13) demonstrated RNA-RNA hybridization between the 5′-region of Ω-U1, containing the AUU sequence, and the 3′-terminus of wheat germ 18S rRNA, again supporting the possibility of disome formation. In addition to TMV, several other viral RNA leader sequences have been shown to form disome (or even trisome) structures (2, 14-16). The 36-nucleotide leader of AlMV RNA4 binds only one ribosome (17), nevertheless it will stimulate expression of contiguous foreign gene transcripts in vitro (18).

We wished to determine whether translational enhancement was a general feature of 5′-untranslated viral leader sequences and

if the ability of a viral leader to form disomes could be correlated with its ability to enhance translation of a contiguous open reading frame. For this purpose, synthetic oligonucleotide sequences derived from the 5′-leaders of TMV (U1 strain; disome), TMV (SPS strain; disome), turnip yellow mosaic virus (TYMV; disome), alfalfa mosaic virus (AlMV) RNA 4 (monosome), brome mosaic virus (BMV) RNA 3 (disome), and the animal retrovirus, Rous sarcoma virus (RSV; disome) were analyzed for their relative abilities to stimulate expression of convenient reporter gene transcripts *in vivo*.

MATERIALS AND METHODS

Bacterial strains, plasmids, enzymes, and media

Escherichia coli strains HB101 and JM101 were obtained from F. Bolivar and J. Messing, respectively. The pSP64 derivatives pJII1, pJII101, pJII2, pJII102 have been described (7). The chloramphenicol acetyltransferase (CAT) reporter gene from Tn9 was obtained from T.J. Close (CSIRO, Canberra, Australia). The β-glucuronidase gene (GUS) and its derivatives were obtained from R. Jefferson and M. Bevan (Plant Breeding Institute, Maris Lane, Trumpington, Cambridge). SP6 RNA-polymerase, human placental RNase inhibitor, DNA polymerase I (Klenow fragment), T4 DNA ligase and all restriction endonucleases were purchased from Boehringer (Mannheim), Pharmacia Ltd., or New England BioLabs. Purified CAT was bought from Pharmacia Ltd. SOC medium (19) was used to prepare competent *E. coli* cells, and L-broth (20) was used for all other cultures.

Plasmid DNA purification and manipulation

Preparative scale (21) and small scale (22) DNA isolations were as described. Standard DNA manipulations were performed essentially as described (21).

Oligodeoxyribonucleotide synthesis

Oligodeoxyribonucleotides were synthesized by S. Gilmore and A.J. Northrop (Institute of Animal Physiology, Babraham, Cambridge) using a Biosearch 8600 4-channel DNA synthesizer and the β-cyanoethyl-phosphoramidite method (23). For each full-length dsDNA viral leader, one complete strand (the coding strand) was synthesized with a 5′-*Hin*dIII site (+1 base) and a

3′- *SalI* site (+ 1 base), for subsequent insertion into the transcription plasmid pJII1. A second complementary oligodeoxyribonucleotide (24-mer) was then annealed, and the dsDNA filled-in by polymerization with either DNA polymerase I (Klenow fragment) or reverse transcriptase.

Construction of trp promoter plasmid pJII168 for E. coli transformation

A 90 base pair (bp) *HindIII/BamHI* fragment containing the tryptophan (*trp*) promoter (P-L Biochemicals, Inc.) was introduced into the *HindIII/BamHI* sites of pJII1 (7), from which a 400bp *BamHI* fragment containing the TMV origin-of-assembly sequence had been removed. The *HindIII* site upstream from the *trp* promoter was removed by digestion with *HindIII*, filling-in with DNA polymerase I (Klenow fragment), followed by re-ligation. A *HindIII* site was then introduced at the 3'-end of the promoter by replacing the 25bp *HpaI/SalI* fragment with a synthetic 17bp *HpaI/SalI* fragment, containing a *HindIII* site positioned at the transcription start site.

RNA synthesis

In vitro transcription of linearized plasmid DNAs was carried out using bacteriophage SP6 RNA polymerase (24). Capped transcripts were obtained by modifying the published reaction conditions to include 200µ*M* GTP and 1.5m*M* $G^{5'}ppp^{5}G$ (Pharmacia, Ltd). RNAs were quantitated either by trace-labelling with α-[^{32}P]-rUTP or by formaldehyde-agarose gel electrophoresis as described (24).

Preparation and electroporation of tobacco mesophyll protoplasts

Mesophyll protoplasts were isolated from leaves of *Nicotiana tabacum* (cv. Xanthi) and stored in 0.7*M* mannitol (25). Electroporation of RNA into protoplasts and incubations were carried out as previously described (7).

After incubation, electroporated protoplasts were sedimented, resuspended and broken by ten passages through a 26-gauge needle in 400µl of 0.25*M* Tris-HCl, pH 7.4, containing 10m*M* dithiothreitol (DTT). Extracts were microcentrifuged at 10,000x*g* for 10 min at 4°C.

Microinjection of Xenopus laevis oocytes

X. laevis were purchased from Xenopus Ltd., South Nuffield,

U.K. Two ng of each synthetic uncapped SP6 mRNA were injected into the cytoplasm of stage 6 oocytes in batches of 25 using standard procedures (26). Oocytes were incubated for 21 hours in Modified Barth's Saline, then washed briefly in distilled water. Extracts from *Xenopus* oocytes were prepared by resuspending each sample in 0.25*M* Tris-HCl, pH 7.4, 10m*M* DTT (20µl/oocyte), followed by sonication for 10 sec. Insoluble material was removed by microcentrifugation for 15 min and fractions of the supernatant representing equivalent numbers of oocytes were assayed for CAT activity.

CAT assay

The protein concentration of each supernatant from *X. laevis* oocytes or tobacco protoplasts was determined by the method of Bradford (27). The CAT assay was essentially as described (28), but used 0.25*M* Tris-HCl, pH 7.4, containing 10m*M* DTT and 30m*M* acetyl-CoA. Quantitation of the thin-layer chromatograph was achieved by cutting out the area corresponding to each ^{14}C-labelled spot and counting in a toluene-based scintillant containing 4% (w/v) PPO and 0.005% (w/v) POPOP.

GUS assay

GUS activity was measured spectrophotometrically or fluorimetrically in 0.5ml assay buffer containing 50m*M* sodium phosphate, pH 7.0, 10m*M* 2-mercaptoethanol, 0.1% (v/v) Triton X-100 and either 1m*M* *p*-nitrophenyl-β-D-glucuronide (PNPG; for *E. coli* extracts) or 0.5m*M* 4-methyl-umbelliferyl-β-D-glucuronide (MUG; for tobacco protoplast extracts). Assays were carried out at 37°C and were terminated by addition of 0.4ml 2.5*M* 2-amino-2-methyl-1,3-propanediol for *E. coli* extracts, or 0.5ml 0.2*M* Na_2CO_3 for protoplast extracts. *p*-Nitrophenol absorbance was measured at 415nm using a Pye Unicam SP1800 Spectrophotometer. Fluorescence was measured by excitation at 365nm and emission at 455nm in a Perkin-Elmer 204 Fluorescence Spectrophotometer.

In situ localization of GUS activity in SDS-polyacrylamide gels

Samples of protoplast extracts containing equivalent amounts of protein were incubated with an equal volume of gel loading buffer (29) at room temperature for 15 min, followed by SDS-polyacrylamide gel electrophoresis (29) in a 12.5% (w/v) gel at 50V for 16 hours. The gel was rinsed 4 times in 100ml assay

buffer (without the glucuronide substrate) for a total of 2 hours, incubated on ice in assay buffer containing 0.5mM MUG for 30 min, and transferred to a glass plate at 37°C for 30 min. The gel was then sprayed with 0.2M Na_2CO_3 and photographed under long-wavelength ultraviolet light using a Wratten 2E filter.

RESULTS

Quantitation of the effect of Ω'-U1 and initiation codon context on expression of GUS mRNAs in tobacco protoplasts.

A derivative of the TMV leader, Ω'-U1 (Fig. 1), has been shown to enhance translation of CAT mRNA in tobacco mesophyll protoplasts, and other eukaryotic and prokarotic systems (7). To quantitate the effect of Ω'-U1 in protoplasts more precisely, we used the GUS reporter gene (30). A SalI-ended fragment containing the GUS gene from pRAJ235 (30) was introduced into the SalI site of the pSP64-derived vectors pJII101 and pJII1 (7), resulting in pJII120 and pJII119, with or without a 5′-proximal Ω'-U1 sequence, respectively. The native SalI GUS fragment had 19 nucleotides upstream of the AUG start codon (Fig. 1). The context of this AUG codon (5′-CCCUUAUGU-3′) was, according to Kozak (6), inefficient for eukaryotic translation (hereafter referred to as "bad context" GUS). To determine whether the effect of Ω'-U1 on mRNA expression was influenced by the context of the initiation codon, a SalI fragment of a derivative of the GUS gene with an initiation codon context (5′-CGACCAUGG-3′) close to the consensus sequence for optimal eukaryotic translation initiation was constructed (in pRAJ275; (8)). This derivative (hereafter referred to as "good context" GUS) had only 7 nucleotides upstream of the AUG (Fig. 1). This SalI fragment was introduced into pJII101 and pJII1 as for "bad context" GUS, resulting in pJII140 and pJII139, with or without a 5′-proximal Ω'-U1 sequence, respectively. 5′-Capped or uncapped mRNAs were synthesized *in vitro* by SP6 RNA polymerase on BglII-linearized pJII119, pJII120, pJII139, pJII140 templates. Eight micrograms of each transcript were electroporated into tobacco mesophyll protoplasts and incubated for 20 hours at 25°C. Assaying protoplast extracts by GUS-activity gel (Fig. 2) revealed that "good context" GUS mRNAs (tracks 6-9) were expressed more efficiently

HindIII Sal I
AAGCTT ▬▬▬ GTCGAC

m^{7}GpppGUAUUUUUACAACAAUUACCAACAACAACAAACAACAAACAACAUUACAAUUACUAUUUACAAUUACAAUGG — TMV-U1 strain

AAGCTT ▬▬▬ GTCGAC

m^{7}GpppGUAUUUUUACAACAAUUACCAACAACAACAACAAACAACAACAACAUUACAUUUUACAUUCUACAACUACAAUGG — TMV-SPS strain

AAGCTT ▬▬▬ GTCGAC

m^{7}GpppGUUUUUAUUUUUAAUUUUCUUUCAAAUACUUCCAUCAUGA — AlMV RNA4-L strain

AAGCTT ▬▬▬ GTCGAC

m^{7}GpppGUAAAAUACCAACUAAUUCUCGUUCGAUUCCGGCGAACAUUCUAUUUUACCAACAUCGGUUUUUUCAGUAGUGAUACUGUUUUUGUUCCCGAUGU — BMV RNA 3

AAGCTT ▬▬▬ GTCGAC

m^{7}GpppGUAAUCAACCACCAAUUCCAGCUCUCUUUUGACAACUGGUCUUAUACCCACUUCCGUACACUUGCAACCCUCGUAAGACAAUUGCAAAUGAGUAAUGG — TYMV RNA (genomic)

AAGCTT ▬▬▬ GTCGAC

m^{7}GpppGmCCAUUUGACCAUUCACCACAUUGGUGUGCGCCUGGGUUGAUGGCCGGACCGAUGAUUCCCUGACGACUACGAGCACAUGCAUGAAGCAGAAGGCUUCAUUUGGUGACCCCGACGUGA
UUGUUAGGGAAUAGUGGAUGGCCAUAGACGGCUCGGAGACCUCGUCCUCACCCGUCUAGCUUAUUCGGGGAGCGGACGAUGACCCUAGUAGAGGGGGCUGCGGCUUAGGAGGGCAGAAGCUGAGU
GGCGUCGGAGGGAGCUCUAUCGUCGGCGAGCUAACAUACCCUACCGAGAACACAGAGAGUCGUUGGAAGUCGGGUAGGUUGCCGACCGACUGAGCGGUCCUCCCCAGGCGAGACCUUGGUCGCUC
GGUGGAUCAAGCAUGG — RSV-Pr-B strain

Sal I
GUCGACGAGCUUUUCAGGAGCUAAGGAAGCUAAAAUGG — CAT Leader

GUCGACCGGUCAGUCCCUUAUGU — GUS Leader "Bad" Context

GUCGACCAUGG — GUS Leader "Good" Context

Fig. 1. DNA constructs representing the 5′-untranslated viral leaders tested for translational enhancement. The sequence of the untranslated portion of each viral RNA up to position +4 (6) of the first open reading frame is shown. Each initiation codon (AUG) is underlined. The region of each leader sequence used in the construction of the corresponding oligodeoxyribonucleotide is marked above the RNA sequence by the bold line (the uppermost being Ω′-U1). Terminal restriction sites for HindIII and SalI were present in each DNA construct. Additional nucleotides present between the SalI site and the start codon (underlined) of the CAT or GUS reporter gene cassettes are shown below as RNA sequences.

than "bad context" GUS mRNAs (tracks 2-5). In addition, the presence of Ω′-U1 on both the "good" and "bad context" GUS mRNAs enhanced expression considerably, whether the mRNAs were capped or not. Accurate fluorimetric quantitation of the kinetics of GUS activity (Fig. 3) and hence of GUS mRNA expression (Table 1), revealed that the levels of expression of uncapped "good" or "bad context" GUS mRNAs were below the limit of detection and only became detectable when the transcripts were capped and/or when Ω′-U1 was present. In all cases, the presence of Ω′-U1

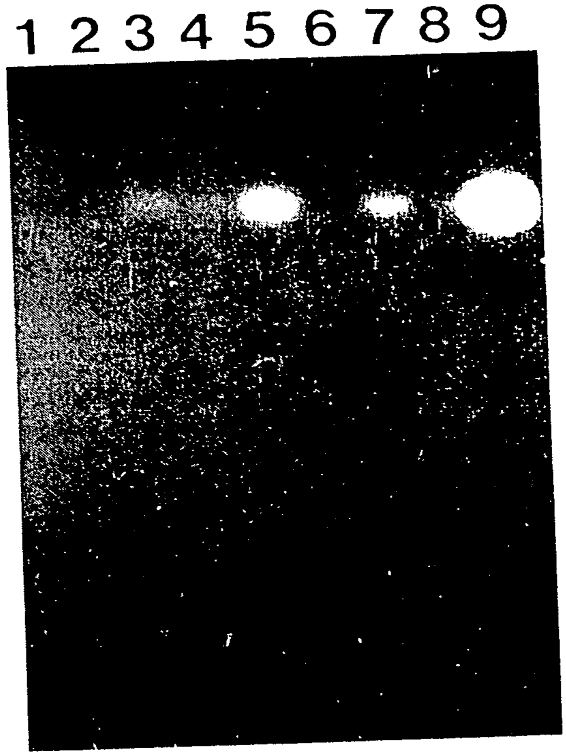


Fig. 2. β-glucuronidase activity-gel of extracts from electroporated tobacco mesophyll protoplasts. Extract volumes representing equivalent amounts of protein were loaded onto each track. Both "bad" and "good context" GUS mRNAs were used to quantitate the effect of Ω'-U1 or a 5′-cap on expression of the enzyme. Electroporated RNAs were: track 1, no RNA (mock); tracks 2-5, "bad context" mRNAs and tracks 6-9, "good context" mRNAs. Tracks 2 and 6, GUS mRNA; tracks 3 and 7, Ω'-U1-GUS mRNA; tracks 4 and 8, 5′-capped-GUS mRNA; tracks 5 and 9, 5′-capped-Ω'-U1-GUS mRNA.

enhanced expression markedly, stimulating the "bad context" GUS mRNA approximately 20-fold. Stimulation of "good context" GUS mRNA by Ω'-U1 was even greater, showing an 80-fold increase with the capped form of the transcript (Table 1).

<u>Other viral leader sequences as translational enhancers in tobacco protoplasts.</u>

To determine whether the phenomenon of translational enhancement is associated with all viral RNA leader sequences or only with those leaders which form disome structures, <u>Hin</u>dIII and <u>Sal</u>I-linkered oligonucleotides were synthesized which incorporated the 5′-leader sequences of: TMV (SPS strain), TYMV, AlMV RNA4, BMV RNA3, and part of RSV RNA (Fig.1). Due to constraints of synthesis only the 5′-112 residues of the 380-nucleotide RSV

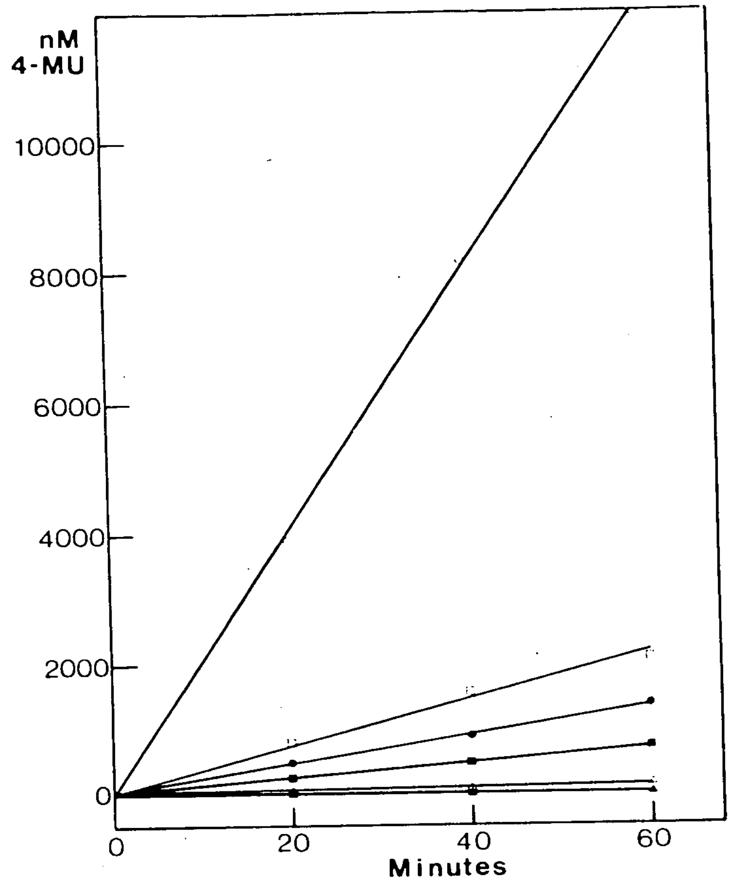


Fig. 3. Kinetic analysis of β-glucuronidase activity in extracts from electroporated tobacco mesophyll protoplasts. Graphical display of the rate of appearance of the reaction product (4-methyl-umbelliferone, 4-MU). Extract volumes representing equivalent amounts of protein were added to each assay. ▲, "bad" or "good context" GUS mRNAs; △, 5′-capped "bad" or "good context" GUS mRNAs; ■, Ω'-U1-"bad context" GUS mRNA; □, Ω'-U1-"good context" GUS mRNA; ●, 5′-capped-Ω'-U1-"bad context" GUS mRNA; ○, 5′-capped-Ω'-U1-"good context" GUS mRNA.

leader (3) were synthesized. This includes the region (residues 9-53) of the native RSV leader shown to act as the binding site for a second 80S ribosome (16). For cloning purposes, these oligonucleotides were manipulated in an identical fashion to Ω'-U1 (7). A family of SP6-transcripts in which each leader was located upstream of the "bad context" GUS gene were electroporated into tobacco protoplasts. Only the Ω'-U1 and, to a lesser extent, the Ω'-SPS leaders proved stimulatory for uncapped transcripts (Table 2). However, when the transcripts were capped, stimulation was observed with the leaders of AlMV RNA4, BMV RNA3

TABLE 1

Translational enhancement by Ω'-U1 on GUS mRNAs electroporated into tobacco protoplasts

SP6-RNAs	Initiation codon context	Specific activity (nmoles MUG hydrolysed/ min/µg protein)	Fold-stimulation
Uncapped			
GUS	bad	<0.01	1
Ω'-U1-GUS	bad	0.18	>18
GUS	good	<0.01	1
Ω'-U1-GUS	good	0.35	>35
5′-Capped			
GUS	bad	0.03	1
Ω'-U1-GUS	bad	0.61	20
GUS	good	0.04	1
Ω'-U1-GUS	good	3.2	80

and "RSV" as well as Ω'-U1 and Ω'-SPS. Only with the TYMV leader did the level of GUS activity remain below the limit of detection.

Other viral leader sequences as translational enhancers in X. laevis oocytes

We have shown (7) that *Xenopus* oocytes, microinjected with capped or uncapped CAT mRNAs, gave approximately 3- to 4-fold more CAT activity when the Ω'-U1 leader sequence was present. In common with most (or all) animal cells, *Xenopus* oocytes contain high levels of endogenous GUS activity. It was therefore not feasible to assay the different viral leader sequences using GUS mRNA as the reporter. Consequently, various pSP64-based leader constructs, each containing the CAT gene, were transcribed and the uncapped mRNAs microinjected into oocytes. In this experiment, the presence of Ω'-U1 gave a 7.5-fold enhancement of CAT activity (Fig. 4). This probably reflects the better quality oocytes than were used previously (7). The Ω'-SPS

TABLE 2

Translational enhancement by various viral leaders on "bad context" GUS mRNAs electroporated into tobacco protoplasts

SP6-RNAs	Specific activity (nmoles MUG hydrolysed/ min/μg protein)	Fold-stimulation
Uncapped		
GUS	< 0.01	1
Ω'-U1-GUS	0.25	> 25
Ω'-SPS-GUS	0.15	> 15
TYMV-GUS	< 0.01	-
AlMV RNA4-GUS	< 0.01	-
BMV RNA3-GUS	< 0.01	-
"RSV"-GUS	< 0.01	-
5′-Capped		
GUS	0.03	1
Ω'-U1-GUS	0.54	18
Ω'-SPS-GUS	0.43	14
TYMV-GUS	< 0.01	-
AlMV RNA4-GUS	0.23	8
BMV RNA3-GUS	0.23	8
"RSV"-GUS	0.23	8

sequence gave a similar (6-fold) level of enhancement. The BMV RNA3, "RSV", and AlMV RNA4 leaders were not stimulatory in this system. The TYMV leader sequence appeared to reduce expression of CAT mRNA.

Enhancement by viral leader sequences in prokaryotic cells

In previous work (7), Ω'-U1 was shown to be stimulatory *in vitro* in an *E. coli* translation system. The reporter gene sequences used encoded CAT or neomycin phosphotransferase (NPTII). In both cases, the transcripts contained the natural prokaryotic

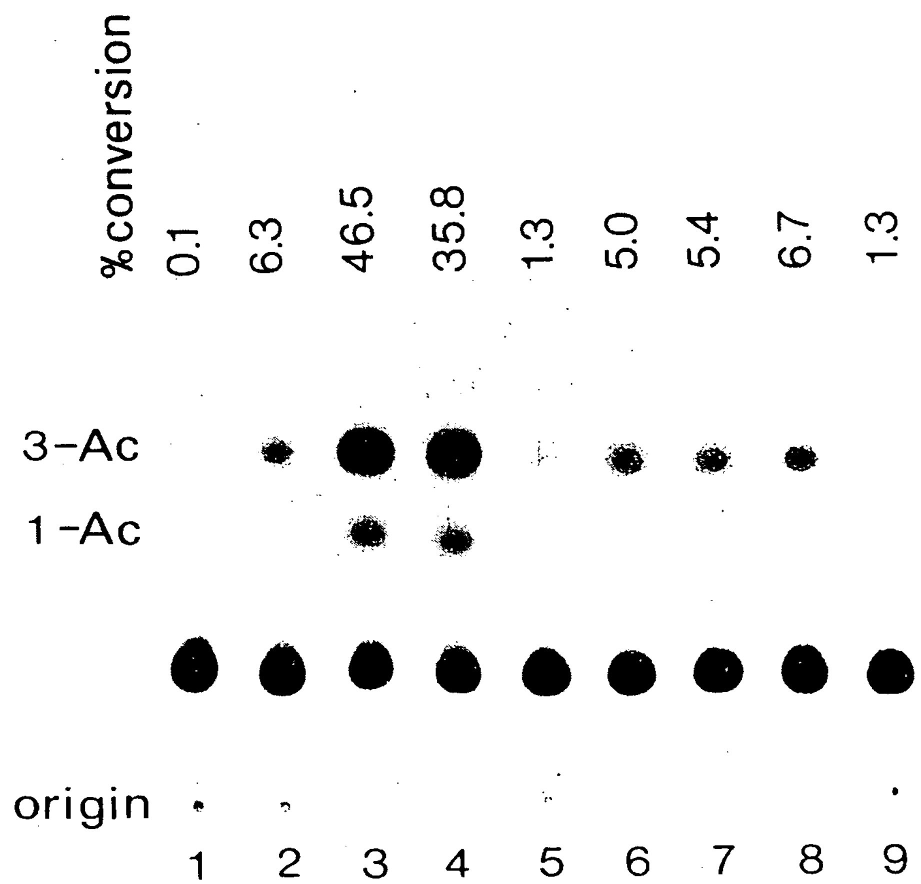


Fig. 4. The effect of various viral leader sequences on expression of CAT mRNAs microinjected into X. laevis oocytes. Oocyte extract volumes (equivalent to 0.25 x cell) were assayed in each case. Conversion (%) of ^{14}C-chloramphenicol into its mono-acetylated form, is shown above each track. Microinjected RNAs were: track 1, no RNA (mock); track 2, CAT mRNA; track 3, Ω'-U1-CAT mRNA; track 4, Ω'-SPS-CAT mRNA; track 5, TYMV-CAT mRNA; track 6, AlMV RNA4-CAT mRNA; track 7, BMV RNA3-CAT mRNA; track 8, "RSV"-CAT mRNA; track 9, 0.1 unit purified CAT enzyme added to an equivalent volume of extract as in track 1. The dried tlc plate was autoradiographed at room temperature for 4 hours before excising and counting the relevent ^{14}C-labelled spots.

Shine-Dalgarno (S-D) ribosome-binding site. The S-D sequence was located between the 3′-end of Ω'-U1 and the start of the open reading frames for CAT or NPTII. The S-D region is considered to be the most critical feature of a prokaryotic mRNA, signalling the attachment of a 30S ribosomal subunit to initiate translation at a downstream start codon. Nevertheless, with Ω'-U1 positioned upstream from the natural S-D region of CAT or NPTII, there was a significant enhancement of translation in vitro in

Operator

-35 -10

CTGTTGACAATTAATCATCGAACTAGTTAACTAGTACGAAGCTTGTCGACGGATCCCCAGATCT

HindIII SalI BamHI BglII

Fig. 5. Trp-promoter construct used to assay the effect of viral leader sequences on expression of various GUS gene transcripts *in situ*, in *E. coli*. -35, -10, and operator regions of the promoter are designated above the sequence. Restriction sites are underlined below. The arrow indicates the site of transcription initiation. GUS gene cassettes were introduced at the SalI site of pJII168 after the various HindIII/ SalI leader cartridges (Fig. 1) had first been inserted.

E. coli. Recently (8), we have shown that Ω'-U1 also stimulates translation of eukaryotic mRNAs, which contain no S-D-like sequence, *in vitro* in an *E. coli* cell-free system.

To complement these observations, we examined the effect of Ω'-U1 on the *in vivo* expression of a prokaryotic mRNA which lacked a S-D region. A derivative of the tryptophan (*trp*) promoter was constructed (Fig. 5) in the plasmid pJII168. Although the HindIII site altered the native sequence of the *trp* operator region slightly, this derivative retained the regulation associated with the wild-type *trp* promoter (data not shown). The position of the HindIII site resulted in addition of only 4 nucleotides upstream of each leader construct, in contrast to the 12-additional nucleotides present in our *in vitro* SP6 transcripts. The SalI-ended "bad context" GUS gene fragment has the native (*E. coli*) context of the AUG codon and 13-nucleotides upstream from the AUG (Fig. 1). In the native GUS gene, the S-D region began just upstream of this 13-nucleotide leader, but this has now been replaced by a sequence containing the SalI site. When the "bad context" GUS gene was introduced downstream of the *trp* promoter and transformed HB101 cells were induced and assayed, a low but measurable level of GUS activity was detected (Table 3). This is in agreement with the previous observation (31) that the presence of a complete S-D sequence can be advantageous but is not essential for gene expression. Insertion of Ω'-U1 between the *trp* promoter and the GUS sequence resulted in a 7-fold increase

TABLE 3
Translational enhancement by various viral leaders on GUS mRNAs in *E. coli* transformed with a recombinant *trp* promoter plasmid

Leader-GUS construct	Initiation codon context	Specific activity (nmoles PNPG converted/ min/μg protein)	Fold-stimulation
GUS	bad	24	1
Ω'-U1-GUS	bad	162	7
GUS	good	3.8	1
Ω'-U1-GUS	good	31	8
Ω'-SPS-GUS	good	7.4	2
TYMV-GUS	good	20	5
AlMV RNA4-GUS	good	21	6
BMV RNA3-GUS	good	10	3
"RSV"-GUS	good	6.4	2

in GUS activity, a level in good agreement with that observed previously for prokaryotic transcripts which contained a S-D region (7,8). This observation contradicts the view that, in all cases, *E. coli* transcripts must have a S-D region for efficient expression. The "good context" GUS mRNA has had the initiation codon context dramatically altered from that of the native gene, and it lacks all the native GUS leader sequence - now replaced with the *Sal*I site and one C-residue (Fig. 1). The *trp* promoter construct containing this "good context" GUS resulted in extremely low, but detectable, levels of GUS activity. Even in this severely altered context, addition of the Ω'-U1 leader produced an 8-fold stimulation in expression of GUS mRNA.

As described above, some residual sequences of the natural GUS mRNA leader were present in the "bad context" GUS construct. Because these might provide some cryptic S-D function, the "good context" GUS construct was chosen as the most sensitive reporter to assay for the effect of the other viral RNA leaders on pro-

karyotic translation *in vivo*. The AlMV RNA4 and TYMV leaders produced a 6- and 5-fold stimulation, respectively (Table 3). The BMV RNA3 and "RSV" leaders provided only slight enhancement, 3- and 2-fold, respectively. Surprisingly, the Ω'-SPS leader sequence was much less stimulatory than Ω'-U1 in *E. coli*, causing only a 2-fold enhancement.

DISCUSSION

Work carried out by Kozak (6) showed that the initiation codon context of eukaryotic mRNAs has an important role in determining the selection of a particular start site and the level of mRNA expression. Our results from protoplasts, using two variants of GUS mRNA with either a "good" or "bad" initiation codon context, support these earlier findings (6).

The endogenous level of GUS activity in tobacco mesophyll protoplasts is extremely low. Thus we were able to quantitate accurately the stimulatory effect of Ω'-U1 and the other viral leaders on expression of GUS mRNA. Whether using "good" or "bad context" GUS mRNA, the presence of Ω'-U1 at the 5′-end resulted in a substantial enhancement of expression (approximately 20-fold; Table 1). When capped mRNAs were used (Table 1), the final level of enhancement by Ω'-U1 with "bad context" GUS mRNA, was greater than 60-fold and, with "good context" GUS mRNA, greater than 320-fold over that seen with the respective GUS mRNAs lacking both a cap and an Ω'-U1 sequence. Similar levels of enhancement were observed with Ω'-SPS (Table 2). In contrast, none of the other viral leader sequences were stimulatory with uncapped GUS mRNAs (Table 2). However, with capped GUS mRNAs, the leader sequences of AlMV RNA4, BMV RNA3, and "RSV" gave a 8-fold enhancement (Table 2). Only the TYMV leader failed to enhance, irrespective of whether the GUS mRNA was capped or not. It is of interest to note that the TYMV leader has been shown to form disomes (14), suggesting that the ability of a leader sequence to form disomes does not correlate with its ability to enhance translation. Alternatively, the 12 additional 5′-nucleotides added by our SP6 vector construct (7), may have selectively destroyed the ability of the TYMV leader to enhance translation. However, it appears that even in the absence of

these additional 5′-nucleotides, the TYMV leader sequence fails to stimulate translation (L. Gehrke, personal communication). *In vivo*, it may be that the TYMV leader is extremely host-dependent in its enhancing ability. Therefore even protoplasts made from tobacco mesophyll cells do not provide the proper machinery for the TYMV sequence. Certainly the ability of a viral leader sequence to enhance translation is not strictly dependent on its capacity to bind more than one ribosome, as shown by data (above and (18)) with the leader of AlMV RNA4 (a monosome former).

Translational enhancement of CAT mRNA by Ω'-U1 in microinjected oocytes was shown previously (7). In this report, enhancement was also observed with the related Ω'-SPS sequence. In contrast, leader sequences from AlMV RNA4, BMV RNA3 and "RSV" failed to enhance translation of CAT mRNA in oocytes. The TYMV leader construct reduced CAT mRNA expression by 80%.

The enhancing effect of Ω'-U1 in *E. coli* cells may be due to some fortuitous interaction with the prokaryotic translation machinery. However, as Ω'-U1 is devoid of G-residues, it cannot provide a sequence similar to that described by Shine and Dalgarno; (5′-AGGAGGU-3′; (1)) and shown to be present in, and required for efficient expression of, nearly all *E. coli* mRNAs studied to date. Of the other viral leaders, only those from TYMV and AlMV RNA4 displayed any significant enhancement of GUS activity in transformed *E. coli* cells.

In this survey of viral RNA leader sequences, only one, TYMV, consistently failed to enhance expression in the plant protoplast system. All the leader sequences were derived from positive-sense RNA viruses which must express their genetic information immediately and efficiently within the infected plant or animal cell to avoid the risk of degradation by host RNases. Furthermore, they must compete effectively with the endogenous cellular mRNAs.

Sequence comparisons between the leaders tested here show no significant homologies other than a high A,U-content, a common feature of viral leader sequences. It is tempting to speculate that the viral leader sequence may circumvent the need for some rate-limiting initiation factor(s), or that it acts as an enhancing element for the association of ribosomes or initiation

factor(s). A precedent for the former possibility exists in the translational regulation of the prokaryotic IF3 gene (32). The sequence immediately surrounding the IF3 start codon allows 30S ribosomal subunits to bind, and translation to begin, without a requirement for IF3, an initiation factor normally essential for the initiation process. The findings of Yokoe and co-workers (13) suggest, at least for the two TMV leaders tested, that a eukaryotic equivalent of the S-D region exists to interact with the 3′-end of 18S rRNA. The lack of homology between the various viral leader sequences may indicate that no one strategy is followed by all, but that there may be several ways to achieve enhancement.

The high A,U-content of these leaders might suggest a low index of secondary structure, which would present fewer obstacles to scanning (4,6) by eukaryotic ribosomes. However, the weak but stable secondary structure potential of the BMV RNA3 leader (2), and the potential of the complete RSV leader to form extensive secondary structures (3), have been used to explain how the 5′-cap and the initiation codon are juxtaposed to facilitate ribosome binding and translation initiation.

The affinity of these sequences for translation initiation factors or other mRNA-binding proteins remains to be tested. Clearly substantial additional work is required to elucidate the mechanism(s) whereby these viral leader sequences can enhance expression of contiguous coding regions in such diverse translation systems.

ACKNOWLEDGEMENTS

We are grateful to Peter Watkins and Clive Mason for excellent technical assistance, Muriel Hobbs for typing the manuscript and staff at the Institute of Animal Physiology, Babraham, Cambridge, U.K., for oligodeoxyribonucleotide synthesis. We also thank Richard Jefferson and Michael Bevan for pRAJ235 and pRAJ275. This work was funded by Diatech Ltd., London. D.E.S. was supported by a CASE studentship from the Science and Engineering Research Council. These results form part of New British Patent Application No. 86 13481.

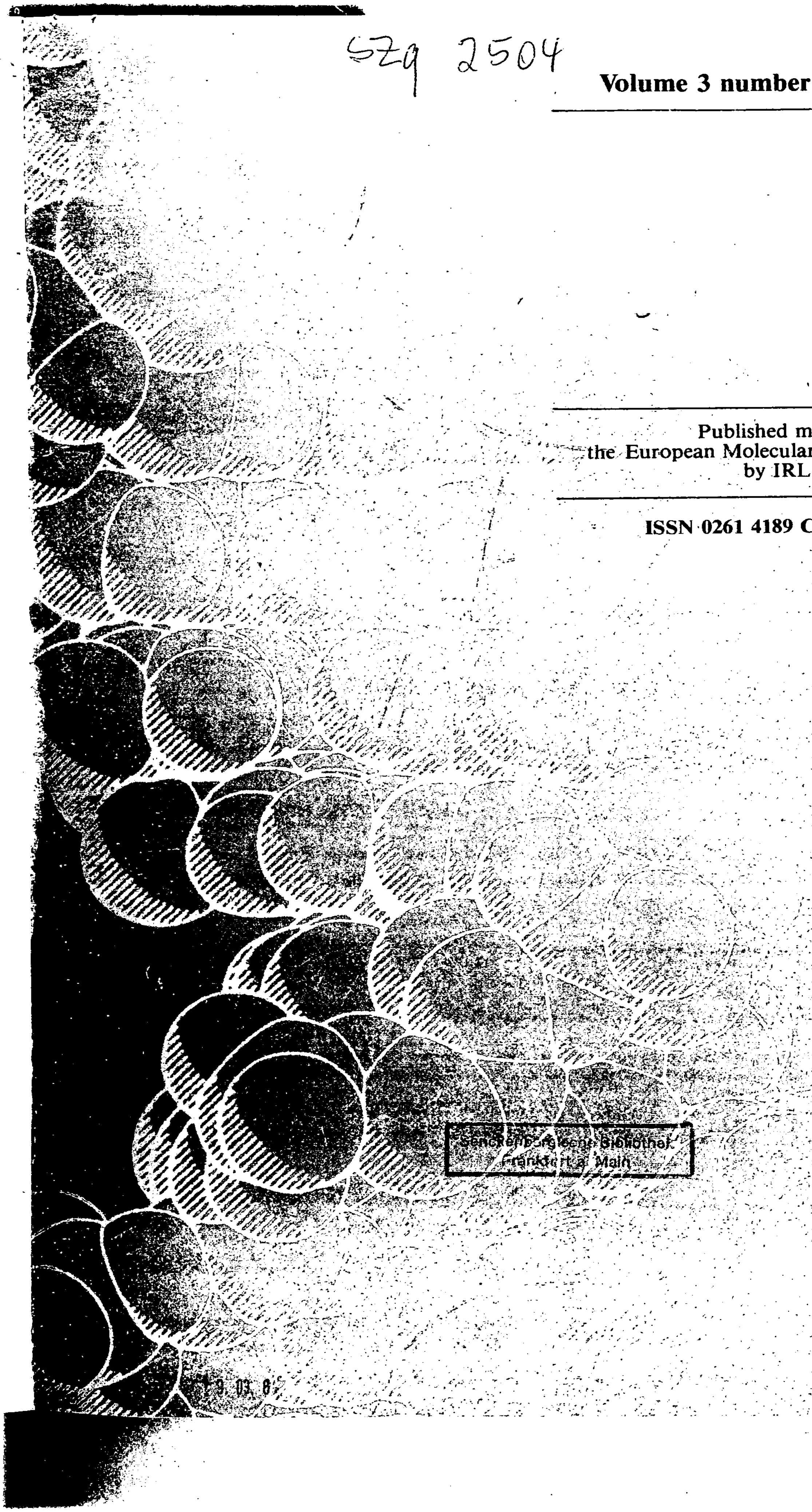
*To whom correspondence should be addressed

REFERENCES

1. Shine, J. and Dalgarno, L. (1975) Nature 254, 34-38.
2. Ahlquist, P., Dasgupta, R., Shih, D.S., Zimmern, D. and Kaesberg, P. (1979). Nature 281, 277-282.
3. Darlix, J-L., Zucker, M. and Spahr, P-F. (1982) Nucl. Acids Res. 10, 5183-5196.
4. Kozak, M. (1986) Proc. Natl. Acad. Sci. USA 83, 2850-2854.
5. Ganoza, M.C., Kofoid, E.C., Marlieve, P. and Louis, B.G. (1987) Nucl. Acids Res. 15, 365-360.
6. Kozak, M. (1986) Cell 44, 283-292.
7. Gallie, D.R., Sleat, D.E., Watts, J.W., Turner, P.C. and Wilson, T.M.A. (1987) Nucl. Acids Res. 15, 3257-3273.
8. Sleat, D.E., Gallie, D.R., Jefferson, R.A., Bevan, M.W., Turner, P.C. and Wilson, T.M.A. (1987), Gene, submitted.
9. Mundry, K.W. (1965) Z. Vererbungsl. 97, 281-296.
10. Mandeles, S. (1968) J. Biol. Chem. 243, 3671-3674.
11. Tyc, K., Konarska, M., Gross, H.J. and Filipowicz, W. (1984) Eur. J. Biochem. 140, 503-511.
12. Hunter, T.R., Hunt, T., Knowland, J. and Zimmern, D. (1976). Nature 260, 759-764.
13. Yokoe, S., Tanaka, M., Hibasami, H., Nagai, J. and Nakashima, K. (1983) J. Biochem. 94, 1803-1808.
14. Filipowicz, W. and Haenni, A-L. (1979) Proc. Natl. Acad. Sci. USA 76, 3111-3115.
15. Pinck, M., Fritsch, C., Ravelonandro, M., Thivent, C. and Pinck L. (1981) Nucl. Acids Res. 9, 1087-1100.
16. Darlix, J-L., Spahr, P-F., Bromley, P.A. and Jaton, J.C. (1979) J. Virol. 29, 597-611.
17. Pinck, L., Franck, A. and Fritsch, C. (1979) Nucl. Acids Res. 7, 151-166.
18. Jobling, S.A. and Gehrke, L. (1987) Nature 325, 622-625.
19. Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.
20. Miller, J.H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. Tait, R.C., Lundquist, R.C. and Kado, C.I. (1982) Mol. Gen. Genet. 186, 10-15.
23. Sinha, N.D., Biernat, J., McManus, J. and Köster, H. (1984) Nucl. Acids Res. 12, 4539-4557.
24. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Nucl. Acids Res. 12, 7035-7056.
25. Motoyoshi, F., Watts, J.W. and Bancroft, J.B. (1974) J. Gen. Virol. 25, 245-256.
26. Colman, A. (1984). In Hames, B.D. and Higgins, S.J. (eds.), Transcription and Translation: A Practical Approach, IRL Press, Oxford, pp. 271-302.

27. Bradford, M.M. (1976) Anal. Biochem. 72, 248-254.
28. Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) Mol. Cell. Biol. 2, 1044-1051.
29. Laemmli, U.K. (1970) Nature 227, 680-685.
30. Jefferson, R.A., Burgess, S.M. and Hirsh, D. (1986) Proc. Natl. Acad. Sci. USA 83, 8447-8451.
31. Ohsawa, H., Herrlich, P. and Gualerzi, C. (1984) Mol. Gen. Genet. 196, 53-58.
32. Gold, L., Stormo, G. and Saunders, R. (1984) Proc. Natl. Acad. Sci. USA 81, 7061-7065.

Volume 3 number 3, March 1984



Published monthly for
the European Molecular Biology Organization
by IRL Press

ISSN 0261 4189 Coden EMJODG

The EMBO Journal vol.3 no.4 pp.835–846, 1984

The complete nucleotide sequence of the TL-DNA of the *Agrobacterium tumefaciens* plasmid pTiAch5

J. Gielen, M. De Beuckeleer[1], J. Seurinck[1,4], F. Deboeck[2], H. De Greve[2], M. Lemmers[1], M. Van Montagu[1,2] and J. Schell[1,3]*

Histologisch Instituut and [1]Laboratorium voor Genetica, Rijksuniversiteit Gent, B-9000 Gent, [2]Laboratorium voor Genetische Virologie, Vrije Universiteit Brussel, B-1640 St.-Genesius-Rode, Belgium, [3]Max-Planck-Institut für Züchtungsforschung, D-5000 Köln 30, FRG
[4]Present address: Plant Genetic Systems, B-9000 Gent, Belgium

*To whom reprint requests should be sent
Communicated by J. Schell

We have determined the complete primary structure (13 637 bp) of the TL-region of *Agrobacterium tumefaciens* octopine plasmid pTiAch5. This sequence comprises two small direct repeats which flank the TL-region at each extremity and are involved in the transfer and/or integration of this DNA segment in plants. TL-DNA specifies eight open-reading frames corresponding to experimentally identified transcripts in crown gall tumor tissue. The eight coding regions are not interrupted by intervening sequences and are separated from each other by AT-rich regions. Potential transcriptional control signals upstream of the 5′ and 3′ ends of all the transcribed regions resemble typical eukaryotic signals: (i) transcriptional initiation signals ('TATA' or Goldberg-Hogness box) are present upstream to the presumed translational start codons; (ii) 'CCAAT' sequences are present upstream of the proposed 'TATA' box; (iii) polyadenylation signals are present in the 3′-untranslated regions. Furthermore, no Shine-Dalgarno sequences are present upstream of the presumed translational start codons.
Key words: Agrobacterium tumefaciens/T-DNA/nucleotide sequence

Introduction

One of the remarkable properties of the Ti plasmids of *Agrobacterium* is their natural capacity to transfer, insert, and express a particular DNA segment of the Ti plasmid in plant cells (for recent reviews, see Nester and Kosuge, 1981; Bevan and Chilton, 1982; Caplan *et al.*, 1983; Zambryski *et al.*, 1983). Depending on the host plant and on the nature of Ti plasmid present in the inciting *Agrobacterium* strain, the transformation event results in crown gall or hairy-root or woolly-knot disease (see Kahl and Schell, 1982).

The segment of Ti plasmid DNA which becomes stably inserted in the plant genome is called T-DNA (Chilton *et al.*, 1977; Lemmers *et al.*, 1980; Thomashow *et al.*, 1980). On the Ti plasmid this DNA segment is bordered by two direct-repeat sequences of 25 bp (Zambryski *et al.*, 1982, 1983; Yadav *et al.*, 1982; Holsters *et al.*, 1983). In the case of the octopine Ti plasmids, two regions of the Ti plasmid, called TL (T-left) and TR (T-right) (Thomashow *et al.*, 1980) according to their position on the standard octopine Ti plasmid map (De Vos *et al.*, 1981) can be transferred and inserted independently into the plant genome. The TL-DNA has been studied more extensively because it encodes essential functions involved in the neoplastic transformation of plant cells (De Beuckeleer *et al.*, 1981; Garfinkel *et al.*, 1981; Leemans *et al.*, 1982; Willmitzer *et al.*, 1982). The TL-DNA also comprises the functions found in common between octopine-type and nopaline-type Ti plasmids' T-regions (Depicker *et al.*, 1978; Chilton *et al.*, 1978; Engler *et al.*, 1981; Willmitzer *et al.*, 1983).

Recently, the nucleotide sequence of the octopine synthase gene (De Greve *et al.*, 1982a), of the gene for 'transcript 7' (Dhaese *et al.*, 1983), and of the gene for 'transcript 4' (Heidekamp *et al.*, 1983) were determined. Here we present the complete nucleotide sequence of the TL-DNA of the *Agrobacterium tumefaciens* plasmid pTiAch5.

Results and Discussion

Sequence determination

To determine the complete sequence of the octopine TL-region, different plasmids containing subfragments of the TL-DNA were constructed (Table I) from clones pGV0153 and pGV0201 (De Vos *et al.*, 1981) containing fragments *Bam*HI-8 and *Hin*dIII-1 (Figure 1), which overlap the complete TL-DNA region. Detailed physical maps of these subclones were established to facilitate the nucleotide sequencing. Plasmid DNA was cleaved with a particular restriction enzyme, and the resulting fragments were ^{32}P end-labeled either at their 5′ termini with polynucleotide kinase or at their 3′ termini with the Klenow fragment of DNA polymerase I. After strand separation or secondary restriction to separate the labeled extremities, the sequence was determined by the limited chemical cleavage method of Maxam and Gilbert (1980). Both DNA strands were sequenced to avoid mistakes that could occur in regions with a distinct secondary structure or by incorrect reading and processing of the sequence information. In addition, as methylated bases (Ohmori *et al.*, 1978) can interfere with correct reading of the sequence, all *Eco*RII sites located in the TL-region were used for sequencing. Furthermore, care was taken that all restriction sites used to generate fragments were resequenced by using another fragment containing an alternative site. Figure 2 gives an overview of the sequence strategy.

Sequence analysis

An uninterrupted sequence of 13 637 bp including the whole TL-DNA of pTiAch5 was determined, and is displayed in the conventional orientation in Figure 3. The numbering starts at the *Hin*dIII site bordering fragments 14 and 18c, which is located 308 bp to the left of the left TL-DNA terminus sequence.

Termini sequences. The TL-region is flanked at both extremities (position 308 and 13 459) by direct repeats of 24 bases, which are believed to be important for the transfer of the TL-DNA segment (Zambryski *et al.*, 1982; Simpson *et al.*, 1982; Holsters *et al.*, 1983).

Table I. Bacterial strains and plasmids

	Antibiotic resistance	Characteristics	Origin
Bacterial strains			
K514		*thr leu thi hsdR*	Colson *et al.* (1965)
SK383	Sm	F⁻ Arg⁻ *his*4, Ilv⁻ *lac*MS286 ϕ80dII*lac*BK1 Sup⁻ *dam*4	S. Kurshner
Plasmids			
pGV0153	Ap	*Bam*HI-8 of pTiAch5 in pBR322	De Vos *et al.* (1981)
pGV117	Ap Cml	*Hind*III-18c of pTiAch5 in pBR325	Dhaese *et al.* (1983)
pGV714	Ap Cml	*Hind*III-22c of pTiAch5 in pBR325	This work
pGV715	Ap Cml	*Hind*III-36 of pTiAch5 in pBR325	This work
pGV716	Ap Cml	*Hind*III-*Bam*HI fragment overlapping the fragments *Bam*HI-8 and *Hind*III-1 in pBR325	This work
pGV0201	Ap	*Hind*III-1 of pTiAch5 in pBR325	De Vos *et al.* (1981)
pGV105	Ap Tc	*Eco*RI-19a of pTiAch5 in pBR325	De Greve *et al.* (1982a)
pGV99	Ap Clm	*Bam*HI-17a of pTiAch5 in pBR325	De Greve *et al.* (1982a)
pGV101	Ap Clm	*Bam*HI-17a of pTiAch5 in pBR325	This work
pGV100	Ap Clm	*Bam*HI-28 of pTiAch5 in pBR325	This work
pGV732	Ap Clm	*Ava*I deletion of pGV101	This work
pGV733	Ap Clm	*Bcl*I deletion of pGV732	This work
pGV734	Ap Clm	*Bcl*I deletion of pGV0201	This work

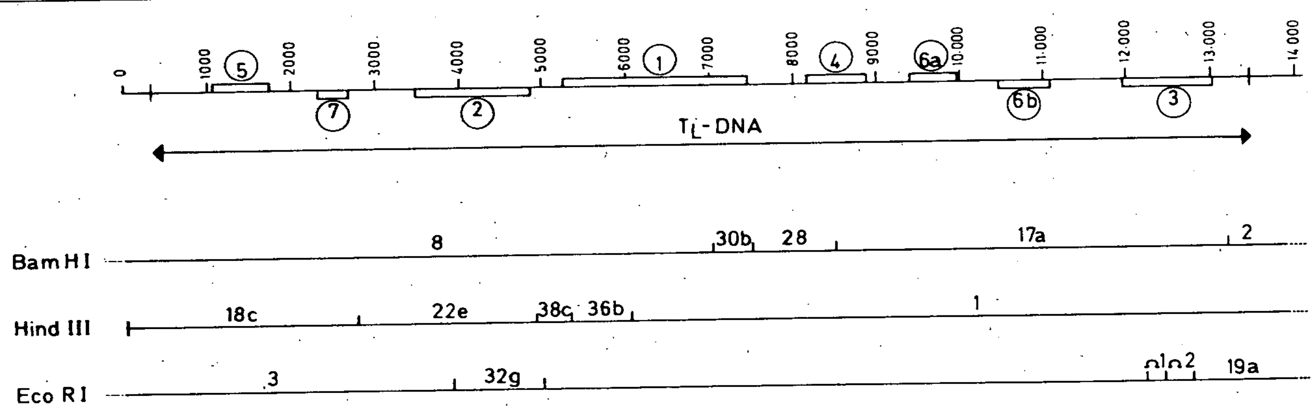


Fig. 1. Restriction map of the TL-DNA of the octopine Ti plasmid pTiAch 5. **Upper portion**: the position of the open-reading frames are presented by open boxes and numbered according to Willmitzer *et al.* (1982). The polarity of the open-reading frames is indicated as follows: open boxes above the line are transcribed from left to right and open boxes below the line are transcribed from right to left. The extent of the TL-DNA is indicated by an arrow and is delimited by the termini boxes (heavy vertical bars). **Lower portion**: a restriction map of the TL-DNA region is shown for the restriction enzymes *Bam*HI, *Hind*III, and *Eco*RI.

A computer search of the complete TL-region for DNA sequences displaying homologies with these direct repeats revealed 10 related DNA sequences. These sequences are listed in Table II. Genetic and physical data indicate that some of these sequences might also be used *in vivo* during transfer and integration of the TL-DNA. Firstly, the sequence (position 11 798) present in the 3′-untranslated region of the octopine synthase gene has been noted by Holsters *et al.* (1983). If this sequence is recognized as a left terminus sequence, the presence of the abbreviated T-DNA found in the octopine-positive regenerate plants rGV1 and rGV5 (De Greve *et al.*, 1982b) can be explained. Alternatively, if this sequence is recognized as a right terminus sequence, instead of the normal terminus sequence, tumor lines containing a shorter TL-DNA which do not synthesize octopine (Thomashow *et al.*, 1980; De Beuckeleer *et al.*, 1981; Ooms *et al.*, 1982) are formed. The origin of teratomas (unpublished results) expressing transcripts 4, 6a, 6b, octopine synthase, and possibly transcript 1, can be explained if the sequence (position 3750) located in transcript 2 is used as a left terminus sequence. Similarly, an abnormal plant (unpublished data) possibly containing transcript 4 and expressing the octopine synthase gene, could be explained if the sequence (position 7777) is used as a left terminus sequence. In addition, either the sequences at position 9078, 10 131, or 10 603 if used as a right terminus sequence, could explain the short TL-DNA observed in a *Petunia* tumor line P-Ach5 (De Beuckeleer *et al.*, 1981). Whether the other sequences also signalled the creation of abbreviated TL-DNAs is difficult to answer because in most cases the resulting transferred DNA

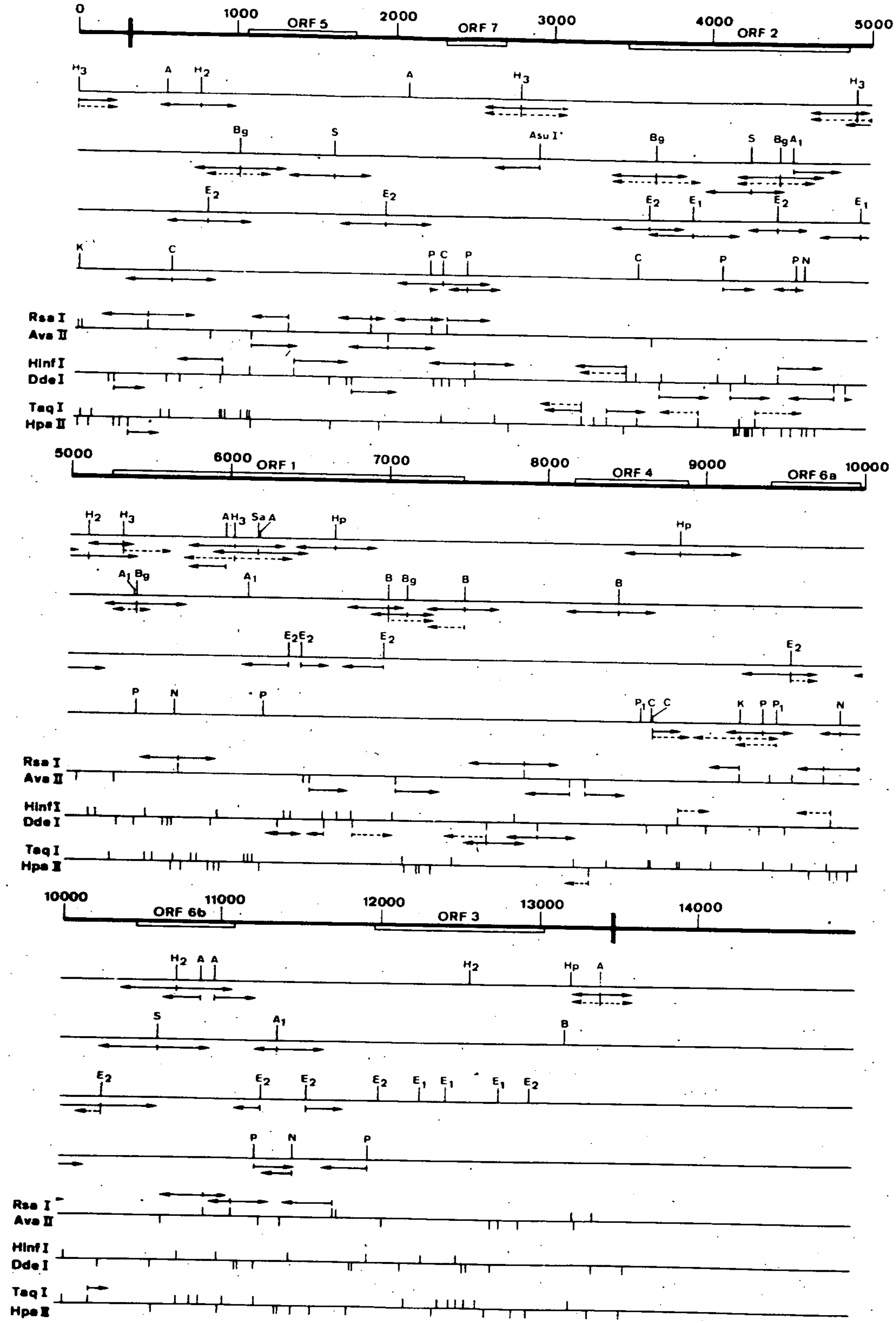


Fig. 2. Sequencing strategy. On a map of the TL-region of pTiAch5 the restriction sites for the following enzymes have been indicated: A, *Acc*I, A1, *Ava*I; Bg, *Bgl*II; C, *Cla*I; E1, *Eco*RI; E2, *Eco*RII; H2, *Hind*II; H3, *Hind*III; Hp, *Hpa*I; K, *Kpn*I; N, *Nae*I; P, *Pvu*II; P1, *Pst*I; S, *Sma*I; Sa, *Sal*I. The position and extent of each sequencing experiment is indicated by a full arrow for a 5′ to 3′ sequencing, and a dashed one for 3′ to 5′. Termini boxes are indicated by a heavy bar, and the open-reading frames corresponding to plant transcripts by open boxes. The polarity of the open-reading frames is indicated from left to right by drawing the open boxes above the line and from right to left by drawing the open boxes below the line.

100
AAGCTTGCTTGGTCGTTCCGGTACCGTGAACGTCGGCTCGATTGTACCTGCGTTCAAATACTTTGCGATCGTGTTGCGCGCCTGCCCGGTGCGTCGGCTGATCTCACGGATCGACTGCTTCTCTCGCAACGCCATCCGACGGATGATGTT
HindIII
200 300
TAAAAGTCCCATGTGGATCACTCCGTTGCCCCGTCGCTCACCGTGTTGGGGGGAAGGTGCACATGGCTCAGTTCTCAATGGAAATTATCTGCCTAACCGGCTCAGTTCTGCGTAGAAACCAACATGCAAGCTCCACCGGGTGCAAAGCGG

400
CAGCGGC $\boxed{\text{GGCAGGATATATTCAATTGTAAAT}}$ GGCTTCATGTCCGGGAAATCTACATGGATCAGCAATGAGTATGATGGTCAATATGGAGAAAAAGAAAGAGTAATTACCAATTTTTTTTCAATTCAAAAATGTAGATGTCCGCAGC
500
GTTATTATAAAATGAAAGTACATTTTGATAAAACGACAAATTACGATCCGTCGTATTTATAGGCGAAAGCAATAAACAAATTATTCTAATTCGGAAATCTTTATTTCGACGTGTCTACATTCACGTCCAAATGGGGGCTTAGATGAGAAA
600 700
CTTCACGATCGATGCCTTGATTTCGCCATTCCCAGATACCCATTTCATCTTCAGATTGGTCTGAGATTATGCGAAAATATACACTCATATACATAAATACTGACAGTTTGAGCTACCAATTCAGTGTAGCCCATTACCTTACATAATTCA
800
CTCAAATGCTAGGCAGTCTGTCAACTCGGCGTCAATTTGTCGGCCACTATACGATAGTTGCGCAAATTTTCAAAGTCCTGGCCTAACATCACACCTCTGTCGGCGGCGGGTCCCATTTGTGATAAATCCACCATCACAATAGATAGTCTA
900 1000
ATGGACGAAAAAGGCGAATATTTCGATGCTGAGATTCGACGCAATTAATTCGAGAAAAATCCCGTGATTGATGCTGTTGAGTTACCAATAATATGGGCAGCGAAGGCCATTTAATTATAAGATCTAACAGAGTTTATATTCAAAAATC

1100
Met Tyr Asp Gly Gln Pro Ile Phe Asn Ile Ile Asp Ser Ser Asn Leu Gln Asp Arg Arg Glu Leu Lys Leu Val Leu Ile His Thr Glu Asn Ala Tyr Arg
AGTCACTAATTCGAT ATG TAT GAC GGT CAG CCG ATA TTC AAC ATT ATC GAC AGC TCG AAT CTA CAG GAC CGG CGT GAA CTT AAA CTC GTC CTA ATT CAC ACA GAG AAT GCT TAT CGC

1200
Ser Ser Ala Gln Arg Ser Leu Ile Ala Ser Gln Arg Ser Trp Ile Asn Phe Ile Ile Asn Thr Asp Val Pro Ile Asp Pro Ala Lys Asp Glu Val Val Lys Cys Ser Arg Lys
AGT TCT GCA CAA AGA AGT CTC ATA GCT TCT CAA CGC TCG TGG ATA AAT TTC ATC ATT AAT ACT GAT GTT CCC ATT GAT CCA GCT AAA GAC GAA GTG GTC AAG TGC TCT CGC AAA

1300
Val Ala Cys Cys Pro Asp Pro Thr Asp Ile Pro Phe Asp Ile Leu Asn Val Ser Leu Leu Tyr Val Tyr Cys Ser Phe Gln Glu Met Arg Arg Tyr Ala Gln Gln Arg Phe Tyr
GTT GCG TGC TGC CCA GAC CCA ACA GAT ATA CCC TTT GAT ATA CTC AAT GTA TCT TTG CTG TAC GTT TAT TGT TCG TTT CAA GAA ATG CGA AGG TAT GCA CAA CAG CGA TTC TAT

1400 1500
Asp Gly Val Ser Asp Gly Gly Ala Val Ile Ser Thr Val Pro Pro Tyr Ala Glu Gly Ile Thr Lys Gln Thr Met Arg Leu Trp Gln Lys Lys Val Trp Gln Asn Thr Ser Lys
GAC GGC GTT TCC GAC GGA GGC GCA GTT ATC TCC ACC GTC CCA CCC TAT GCC GAA GGA ATA ACA AAA CAA ACT ATG AGG TTG TGG CAG AAA AAA GTT TGG CAA AAT ACA AGC AAA

1600
Glu Thr His Asp Leu Asp Ala Tyr Ile Ala Leu Leu Pro Asn Ala Ser Leu Gln Asn Pro Asn Phe Ser His Met Lys Ile Gly Gly Asn Ser Phe Leu Ala Pro Ser Arg Val
GAA ACA CAT GAT TTG GAT GCT TAC ATT GCT CTT CTT CCG AAC GCT TCG CTT CAA AAT CCA AAT TTC TCA CAT ATG AAG ATC GGC GGC AAT AGC TTC TTA GCG CCA TCC CGG GTT
SmaI

1700
Asp Pro Ile Cys Val Glu Ile Val Ala Val Gly Lys Ala Leu Phe Gln Lys Asp Arg Arg Pro Lys Glu Pro Lys Val Arg Trp Ala Met Ala Leu Ser Ser Leu Trp Lys Arg
GAT CCT ATC TGT GTT GAA ATA GTT GCG GTG GGC AAG GCT CTC TTT CAG AAA GAC AGG CGG CCA AAG GAA CCC AAG GTG AGG TGG GCT ATG GCT CTC AGT TCC TTG TGG AAG CGC

Leu Val
1800
TTG GTC TAA GGTGCAGAGGTGTTAGCGGGGATGAAGCAAAAGTGTCCGATTGTAACAAGATATGTTGATCCTACGTAAGGATATTAAAGTATGTATTCATCATTAATATAATCAGTGTATTCCAATATGTACTACGATTTCCAATGT
1900 2000
CTTTATTGTCGCCGTATGTAATCGGCGTCACAAAATAATCCCCGGTGACTTTCTTTTAATCCAGGATGAAATAATATGTTATTATAATTTTTGCGATTTGGTCCGTTATAGGAATTGAAGTGTGCTTGAGCTCGGTCGCCACCACTCCCA
2100
TTTCATAATTTTACATGTATTTGAAAAATAAAAATTTATGGTATTCAATTTAAACACGTATACTTGTAAAGAATGATATCTTGAAAGAAATATAGTTTAAATATTTATTGATAAAATAACAAGTCAGGTATTATAGTCCAAGCAAAAACAT
2200 2300
AAATTTATTGATGCAAGTTTAAATTCAGAAATATTTCAATAACTGATTATATCAGCTGGTACATTGCCGTAGATGAAAGACTGAGTGCGATATTATGTGTAATACATAAATTGATGATATAGCTAGC TTA GCT CAT CGA TCC ATG
Ser Met Ser Gly His

2400
GGC TAC TAT GGG GTA CAG AAA TGG GCG ATT ATG GCA TCT CAG AAA GCC TTT CTC TGG ATT TGA AAC GCA CAG GAA ATA GTT GCT TTG AAA AAT GGC GAC ATA ATA AGT TAA ATC
Ala Val Ile Pro Tyr Leu Phe Pro Arg Asn His Cys Arg Leu Phe Gly Lys Glu Pro Asn Ser Val Cys Leu Phe Tyr Asn Ser Gln Phe Ile Ala Val Tyr Tyr Thr Leu Asp

2500
ACT TTG TTG CGC CTT CAT CTC AGC TGG CTT TAT GGT AGT GAA GGA TAA TTC TTC CTT CTC CTT AAA TTT GAG GTG TGT GTC ATG AAT CCG CTG TGA GAG TGA ACC TTT GCC ATA
Ser Gln Gln Ala Lys Met Glu Ala Pro Lys Ile Thr Thr Phe Ser Leu Glu Glu Lys Glu Lys Phe Lys Leu His Thr Asp His Ile Arg Gln Ser Leu Ser Gly Lys Gly Tyr

2600
CAA ATA CAG CAG CCC ATT GTT TGT TTG GAT TAC CTC TCC TGT TTC CAA TTG TGG AGA TGC ACC ATA AGT TTT GAT AAA CTC TTC GCA TGC CCA GTC TAG GTC GAG GGA GGC CAA
Leu Tyr Leu Leu Gly Asn Asn Thr Gln Ile Val Glu Gly Thr Glu Leu Gln Pro Ser Ala Gly Tyr Thr Lys Ile Phe Glu Glu Cys Ala Trp Asp Leu Asp Leu Ser Ala Leu

2700 2800
GGG AGT ATC TGC GAA ATT CAT GATGTGAGGTGTGTTTGTGTAAGACAAACTATATTGGCGCTATATATAGATGCCGGGCTGGGACGACCTGTGGCGATGATTTCGGGGGGGGGGCTTAATTTTCAAGCTTGAACTAGTCTCA
Pro Thr Asp Ala Phe Asn Met HindIII

2900
TCTGAAGAAAATCCATAAAATGCTCGTCACGCTTGGCGGATTTTGCGCTGGCTGCAATTGCTGACTTTTTGTAAGTCTGCGTTTATCTGTGTGCCAAGTGGGCCGCCGACAGCGTAACGTCAATGGCTGACAATGATGATGTTGTAGCA
3000 3100
GGTGAGAGGTGCATCCAAATTAAAAGGTGGGTGCCTTCACGTCTGTCCTCACACGGCGAGACAATTCAAAAAAGTCATTAATTTCATAATGCAGATTTGACAAATTTGTAAAGGATAGTGGACGGCAAATTATATATTATATAAATTGT
3200
ATTAACTACTTTATGCCTAAATAGGATTGCTTGAACTTTAATTATATTTCCCTATAATTTAGGAAAAATGTAATTTGCTTAAGATATATAATTTAAATTCGAAAAGAATTATTGTTATATGAAGATGCCTATTCCAAGAAATATCTAAA
3300 3400
GTTTATCACTGATAATAAAATTATTTATCGAACATGATTATTGCAAAGACTTTTATTGGTTAAATCATAAATTAAAGTTTGTTCAAAATCTCCATCGCAAGTTATTATTCGATTTATAACATAAGGCACTGTTGTTCATAGACGCAATA

3500
CAAGGATTGATTGTCATCAATCTGAAAAATTGTAAAAACGAACATGGTAGAAAGT TTA ATT GGG TAA ACC GGC AAA ATA TCG GAA TCC AAT GGC TTC TTC CAA TGC CCC CCC GAT TGC TAA CAG ACG
Asn Pro Leu Gly Ala Phe Tyr Arg Phe Gly Ile Ala Glu Glu Leu Ala Gly Gly Ile Ala Leu Leu Arg

3600
TTG GTC TGA ATC CGC TAA TCC ATC GAT CTC CAT TCC AAC AGG CAA GCG ATC AGG TGT CAG GCA AAC AGG AAT GCT CAA GCC AGG TAG GCC TGC GTT GCT GCT TGG GTC CAC ATT
Gln Asp Ser Asp Ala Leu Gly Asp Ile Glu Met Gly Val Pro Leu Arg Asp Pro Thr Leu Cys Val Pro Ile Ser Leu Gly Pro Leu Gly Ala Asn Ser Ser Pro Asp Val Asn

3700
TCG CAC GTA GAT CTT GAA TGT GTC CAG CAT CGT GCC ATT GTG GAT AAC TGA GGA ATC CTG ACC TAT GGG TCT GGC CAC CAA GGG TGC TGT TGG GAA GAG AAT AGC ATC TAA TCT
Arg Val Tyr Ile Lys Phe Thr Asp Leu Met Thr Gly Asn His Ile Val Ser Ser Asp Gln Gly Ile Pro Arg Ala Val Leu Pro Ala Thr Pro Phe Leu Ile Ala Asp Leu Arg

3800
ATT CAG TTT GAA GTA GTT GCG ATA GGT GGC TTG AAG TCT TGG TCT GAA GGA GTG GCG GGC CAG TTC ATA TTC AGC TTT GGA AAT TTG ATG TCC ATC AAT TTG CGC ATT GGC AAT
Asn Leu Lys Phe Tyr Asn Arg Tyr Thr Ala Gln Leu Arg Pro Arg Phe Ser His Arg Ala Leu Glu Tyr Glu Ala Lys Ser Ile Gln His Gly Asp Ile Gln Ala Asn Ala Ile

3900
GTT GGC TAC ATC AGG GCT ACG AAT TCC TTT GAT GAC GTC AGA AAA AGA AAC AGT TTT TAC AAA GTC GTC GAG ATA CTG TTT TAG AGC GTG TGG AAA TTC ATA GAG TGC AAC TGG
Asn Ala Val Asp Pro Ser Arg Ile Gly Lys Ile Val Asp Ser Phe Ser Val Thr Lys Val Phe Asp Asp Leu Tyr Gln Lys Leu Ala His Pro Phe Glu Tyr Leu Ala Val Pro

4000 4100
GAA GCT GGC CCC TTT ATT CAG TTC GTC AAG GTG GGG AAT GTT AGC TTC AAC AAA AGT TAC GCC TTT GTT TGC TAG CAG GCG AAT CGT TGT TTC AGC TGC TAG GGC CAC ATC AGC
Phe Ser Ala Gly Lys Asn Leu Glu Asp Leu His Pro Ile Asn Ala Glu Val Phe Thr Val Gly Lys Asn Ala Leu Leu Arg Ile Thr Thr Glu Ala Ala Leu Ala Val Asp Ala

4200
ATC AAG GTC ATC ATA AAA GTA GGT TGT AGG GAG GCC GAT CCT TAG CCC CTT CAG CGG CAC GGG TGG TAT TCT CTC CGG TGT GCC GGA AAT TAT CCG GTC GAG GAT TAC AAC ATC
Asp Leu Asp Asp Tyr Phe Tyr Thr Thr Pro Leu Gly Ile Arg Leu Gly Lys Leu Pro Val Pro Pro Ile Arg Arg Glu Pro Thr Gly Ser Ile Ile Arg Asp Leu Val Val Asp

4300
GGC TAC GCA CTG CGC TAT GAT TCC GGG AGT GTC CCG GGT AGG GCT AAC CGG TAT TAT CCG ATC TCC CGG ATA TCT ACC AAG CGT CGG TCG AAA TCC TAC TAC GCC ACA CAG GGC
Ala Val Cys Gln Ala Ile Ile Gly Pro Thr Asp Arg Thr Pro Ser Val Pro Ile Ile Arg Asp Gly Pro Tyr Arg Gly Leu Thr Pro Arg Phe Gly Val Val Gly Cys Leu Ala
SmaI

4400
TGC GGG TAG GCG AAC AGA TGC ACC GGT ATC GGT GCC TAT GCC GCC TAA CAT CAA TCG GCT TGC TAC CGC AGC AGC CAC ACC ACC GCT TGA GCC CCC TGG TAT CAG ATC TGG ATT
Ala Pro Leu Arg Val Ser Ala Gly Thr Asp Thr Gly Ile Gly Gly Leu Met Leu Arg Ser Ala Val Ala Ala Ala Val Gly Gly Ser Ser Gly Gly Pro Ile Leu Asp Pro Asn

4500
CCA CGG GTT TCG CAC CGC CCC GGT GGC ATA GTT GTT GCT TGT AAT TCC AAA CGA TAA CTC ATG CAT ATT TCC CGA GGC ACC CGG CAG TGC TCC AGC TGA AAA AAG TCT TTC TGC
Trp Pro Asn Arg Val Ala Gly Thr Ala Tyr Asn Asn Ser Thr Ile Gly Phe Ser Leu Glu His Met Asn Gly Ser Ala Gly Pro Leu Ala Gly Ala Ser Phe Leu Arg Glu Ala

4600
GAC GCG GGA TGG TAT CTT TGG CAA GTG GTT TAT CAG CGC CGG CGT AGC GGC GCT TGT GGG AAA TAC GCC GGT AGC GAT GTT CGC CTT AAA ACA GAG TGG AAT GCC GCA AAG ACC
Val Arg Ser Pro Ile Lys Pro Leu His Asn Ile Leu Ala Pro Thr Ala Ala Ser Thr Pro Phe Val Gly Thr Ala Ile Asn Ala Lys Phe Cys Leu Pro Ile Gly Cys Leu Gly

4700
TAC TCC GGC GTT TCC ATG GCG ATC AAT TTT TTT GGC GCT TCG CCG CAA ACC ATC CCA GTC TGT AGC CAG AAG GGC GTT TAA TGA TTT TGC AGC TTC ACA ACG CGC TAT CAG AGT
Val Gly Ala Asn Gly His Arg Asp Ile Lys Lys Ala Ser Arg Arg Leu Gly Asp Trp Asp Thr Ala Leu Leu Ala Asn Leu Ser Lys Ala Ala Glu Cys Arg Ala Ile Leu Thr

Figure 3(i)

```
                        4800
TTC TAC TAG TTC TAA GCA GGA GTA GTC TTT CCG TTT CAG GTG TTC TAG GCT TTG GGC TAA CGA GGT AAT GGC CAC CAT CTCTCTGAGTTGGAAATTTCAAACCCATTCAGACCAAATAAATATA 4900
Glu Val Leu Glu Leu Cys Ser Tyr Asp Lys Arg Lys Leu His Glu Leu Ser Gln Ala Leu Ser Thr Ile Ala Val Met

CCAAGCTTGAACCCAGCTTGCGCCATTATTGGTGGGAATTCAGCAATCACGCCGTGTGATACGAATGACTGTAGGCACTAGGGTCGCTTTTTCAAACGCAAACAGTATTGTATCAATTATCAATCAATATTCAACTTGATTGAGCGGC 5000
 HindIII
GAGGTGGACCCGCATGAACATCATATTAAAGAAAGCCCAAATTGTCGCTTTCCATTGGCATGTCAACGAACAGTATGTTCCCCGATTCTCAACTTAAATAGAAACGGTTGATGTGGTTATTTATCTACACAAAAACGAATCTTTTCTA 5100 5200

    Met Ser Ala Ser Pro Leu Leu Asp Asn Gln Cys Asp His Leu Pro Thr Lys Met Val Asp Leu Thr Met Val Asp Lys Ala Asp Glu Leu Asp Arg Arg Val Ser Asp Ala
ACA ATG TCA GCT TCA CCT CTC CTT GAT AAC CAG TGC GAT CAT CTC CCA ACC AAA ATG GTG GAT CTG ACA ATG GTC GAT AAG GCG GAT GAA TTG GAC CGC AGG GTT TCC GAT GCC 5300
Phe Leu Glu Arg Glu Ala Ser Arg Gly Arg Arg Ile Thr Gln Ile Ser Thr Glu Cys Ser Ala Gly Leu Ala Cys Lys Arg Leu Ala Asp Gly Arg Phe Pro Glu Ile Ser Ala
TTC TTA GAA CGA GAA GCT TCT AGG GGA AGG AGG ATT ACT CAA ATC TCC ACC GAG TGC AGC GCT GGG TTA GCT TGC AAA AGG CTG GCC GAT GGT CGC TTC CCC GAG ATC TCA GCT 5400
                HindIII
Gly Gly Lys Val Ala Val Leu Ser Ala Tyr Ile Tyr Ile Gly Lys Glu Ile Leu Gly Arg Ile Leu Glu Ser Lys Pro Trp Ala Arg Ala Thr Val Ser Gly Leu Val Ala Ile
GGT GGA AAG GTA GCA GTT CTC TCC GCT TAT ATC TAT ATT GGC AAA GAA ATT CTG GGG CGG ATA CTT GAA TCG AAA CCT TGG GCG CGG GCA ACA GTG AGT GGT CTC GTT GCC ATC 5500
Asp Leu Ala Pro Phe Cys Met Asp Phe Ser Glu Ala Gln Leu Ile Gln Ala Leu Phe Leu Leu Ser Gly Lys Arg Cys Ala Pro Ile Asp Leu Ser His Phe Val Ala Ile Ser
GAC TTG GCA CCA TTT TGC ATG GAT TTC TCC GAA GCA CAA CTA ATC CAA GCC CTG TTT TTG CTG AGC GGT AAA AGA TGT GCA CCG ATT GAT CTT AGT CAT TTC GTG GCC ATT TCA 5600
Ile Ser Lys Thr Ala Gly Phe Arg Thr Leu Pro Met Pro Leu Tyr Glu Asn Gly Thr Met Lys Cys Val Thr Gly Phe Thr Ile Thr Leu Glu Gly Ala Val Pro Phe Asp Met
ATC TCT AAG ACT GCC GGC TTT CGA ACC CTG CCA ATG CCG CTG TAC GAG AAT GGC ACG ATG AAA TGC GTT ACC GGG TTT ACC ATA ACC CTT GAA GGG GCC GTG CCA TTT GAC ATG 5700
Val Ala Tyr Gly Arg Asn Leu Met Leu Lys Gly Ser Ala Gly Ser Phe Pro Thr Ile Asp Leu Leu Tyr Asp Tyr Arg Pro Phe Phe Asp Gln Cys Ser Asp Ser Gly Arg Ile
GTA GCT TAT GGT CGA AAC CTG ATG CTG AAG GGT TCG GCA GGT TCC TTT CCA ACA ATC GAC TTG CTC TAC GAC TAC AGA CCG TTT TTT GAC CAA TGT TCC GAT AGT GGA CGG ATC 5800
Gly Phe Phe Pro Glu Asp Val Pro Lys Pro Lys Val Ala Val Ile Gly Ala Gly Ile Ser Gly Leu Val Val Ala Asn Glu Leu Leu His Ala Gly Val Asp Asp Val Thr Ile
GGC TTC TTT CCG GAG GAT GTT CCT AAG CCG AAA GTG GCG GTC ATT GGC GCT GGC ATT TCC GGA CTC GTG GTG GCA AAC GAA CTG CTT CAT GCT GGG GTA GAC GAT GTT ACA ATA 5900 6000
Tyr Glu Ala Ser Asp Arg Val Gly Gly Lys Leu Trp Ser His Ala Phe Arg Asp Ala Pro Ser Val Val Ala Glu Met Gly Ala Met Arg Phe Pro Pro Ala Ala Phe Cys Leu
TAT GAA GCA AGT GAT CGT GTT GGA GGC AAG CTT TGG TCA CAT GCT TTC AGG GAC GCT CCT AGT GTC GTG GCC GAA ATG GGG GCG ATG CGA TTT CCT CCT GCT GCA TTC TGC TTG 6100
                                    HindIII
Phe Phe Phe Leu Glu Arg Tyr Gly Leu Ser Ser Met Arg Pro Phe Pro Asn Pro Gly Thr Val Asp Thr Tyr Leu Val Tyr Gln Gly Val Gln Tyr Met Trp Lys Ala Gly Gln
TTT TTC TTC CTC GAG CGT TAC GGC CTG TCT TCG ATG AGG CCG TTC CCA AAT CCC GGC ACA GTC GAC ACT TAC TTG GTC TAC CAA GGC GTC CAA TAC ATG TGG AAA GCC GGG CAG 6200
Leu Pro Pro Lys Leu Phe His Arg Val Tyr Asn Gly Trp Arg Ala Phe Leu Lys Asp Gly Phe Tyr Glu Arg Asp Ile Val Leu Ala Ser Pro Val Ala Ile Thr Gln Ala Leu
CTG CCA CCG AAG CTG TTC CAT CGC GTT TAC AAC GGT TGG CGT GCG TTC TTG AAG GAC GGT TTC TAT GAG CGA GAT ATT GTG TTG GCT TCG CCT GTC GCT ATT ACT CAG GCC TTG 6300
Lys Ser Gly Asp Ile Arg Trp Ala His Asp Ser Trp Gln Ile Trp Leu Asn Arg Phe Gly Arg Glu Ser Phe Ser Ser Gly Ile Glu Arg Ile Phe Leu Gly Thr His Pro Pro
AAA TCA GGA GAC ATT AGG TGG GCT CAT GAC TCC TGG CAA ATT TGG CTG AAC CGT TTC GGG AGG GAG TCC TTC TCT TCA GGG ATA GAG AGG ATC TTT CTG GGC ACA CAT CCT CCT 6400
Gly Gly Glu Thr Trp Ser Phe Pro His Asp Trp Asp Leu Phe Lys Leu Met Gly Ile Gly Ser Gly Gly Phe Gly Pro Val Phe Glu Ser Gly Phe Ile Glu Ile Leu Arg Leu
GGT GGT GAA ACA TGG AGT TTT CCT CAT GAT TGG GAC CTA TTC AAG CTA ATG GGA ATA GGA TCT GGC GGG TTT GGT CCA GTT TTT GAA AGC GGG TTT ATT GAG ATC CTC CGC TTG 6500
Val Ile Asn Gly Tyr Glu Glu Asn Gln Arg Met Cys Pro Glu Gly Ile Ser Glu Leu Pro Arg Arg Ile Ala Ser Glu Val Val Asn Gly Val Ser Val Ser Gln Arg Ile Cys
GTC ATC AAC GGA TAT GAA GAA AAT CAG CGG ATG TGC CCT GAA GGA ATC TCA GAA CTT CCA CGT CGG ATC GCA TCT GAA GTG GTT AAC GGT GTG TCT GTG AGC CAG CGC ATA TGC 6600
His Val Gln Val Arg Ala Ile Gln Lys Glu Lys Thr Lys Ile Lys Ile Arg Leu Lys Ser Gly Ile Ser Glu Leu Tyr Asp Lys Val Val Val Thr Ser Gly Leu Ala Asn Ile
CAT GTT CAA GTC AGG GCG ATT CAG AAG GAA AAG ACA AAA ATA AAG ATA AGG CTT AAG AGC GGG ATA TCT GAA CTT TAT GAT AAG GTG GTG GTC ACA TCT GGA CTC GCA AAT ATC 6700 6800
Gln Leu Arg His Cys Leu Thr Cys Asp Thr Asn Ile Phe Gln Ala Pro Val Asn Gln Ala Val Asp Asn Ser His Met Thr Gly Ser Ser Lys Leu Phe Leu Met Thr Glu Arg
CAA CTC AGG CAT TGC CTG ACA TGC GAT ACC AAT ATT TTT CAG GCA CCA GTG AAC CAA GCG GTT GAT AAC AGC CAT ATG ACA GGA TCG TCA AAA CTC TTC CTG ATG ACT GAA CGA 6900
Lys Phe Trp Leu Asp His Ile Leu Pro Ser Cys Val Leu Met Asp Gly Ile Ala Lys Ala Val Tyr Cys Leu Asp Tyr Glu Pro Gln Asp Pro Asn Gly Lys Gly Leu Val Leu
AAA TTC TGG TTA GAC CAT ATC CTC CCG TCT TGT GTC CTC ATG GAC GGG ATC GCA AAA GCA GTG TAT TGC CTG GAC TAT GAG CCG CAG GAT CCG AAT GGT AAA GGT CTA GTG CTC 7000
                                                                                                                              BamHI
Ile Ser Tyr Thr Trp Glu Asp Asp Ser His Lys Leu Leu Ala Val Pro Asp Lys Lys Glu Arg Leu Cys Leu Leu Arg Asp Ala Ile Ser Arg Ser Phe Pro Ala Phe Ala Gln
ATC AGT TAT ACA TGG GAG GAC GAC TCC CAC AAG CTG TTG GCG GTC CCC GAC AAA AAA GAG CGA TTA TGT CTG CTG CGG GAC GCA ATT TCG AGA TCT TTC CCG GCG TTT GCC CAG 7100
His Leu Phe Pro Ala Cys Ala Asp Tyr Asp Gln Asn Val Ile Gln His Asp Trp Leu Thr Asp Glu Asn Ala Gly Gly Ala Phe Lys Leu Asn Arg Arg Gly Glu Asp Phe Tyr
CAC CTA TTT CCT GCC TGC GCT GAT TAC GAC CAA AAT GTT ATT CAA CAT GAT TGG CTT ACA GAC GAG AAT GCC GGG GGA GCT TTC AAA CTC AAC CGG CGT GGT GAG GAT TTT TAT 7200
Ser Glu Glu Leu Phe Phe Gln Ala Leu Asp Thr Ala Asn Asp Thr Gly Val Tyr Leu Ala Gly Cys Ser Cys Ser Phe Thr Gly Gly Trp Val Glu Gly Ala Asn Arg Thr Pro
TCT GAA GAA CTT TTC TTT CAA GCA CTG GAC ACG GCT AAT GAT ACC GGA GTT TAC TTG GCG GGT TGC AGT TGT TCC TTC ACA GGT GGA TGG GTG GAG GGT GCT AAT CGG ACG CCG 7300
Cys Asn Ala Val Cys Ala Ile Ile His Asn Cys Gly Gly Ile Leu Ala Lys Gly Asn Pro Leu Glu His Ser Trp Lys Arg Tyr Asn Tyr Arg Thr Arg Asn
TGT AAC GCC GTC TGT GCA ATT ATC CAC AAT TGT GGA GGC ATT TTG GCA AAG GGC AAT CCT CTC GAA CAC TCT TGG AAG AGA TAT AAC TAC CGC ACT AGA AAT TAG TCTATGGATCC 7400
                                                                                                                                                  BamHI
TGTTACAAGTATTGCAAGTTTTATAAATTGCATATTAATGCAATCTTGATTTTTAACAACGAACGTAATGGCGTAAAAAATGTATGTTATATTATTTATATTTAATTATATTGGAGTGCGCCATAATATGATGATTTATAATTAAAAAATA 7500 7600
TTTACTGTCACATTGACTGAGATGGCACTGTTATTTCAACCATGAAATTTTGTTGATTTTTTTACAATAACAATAATTGCAGGAAGTAAATAATAGACGCCGTTGTTAAAAAATTGCAATCATATGTGCCTAACTATAGGGACAATTAGGT 7700
CAATTGTAATAGTCTCCCTTATTTTAACGACTCACCTAATCAAGTATTACAAAATATCTCACTTTTCGTCAGTAATGATGTAATCAGAACTGAATAGTACAAGTAAAACGTGGAAAAACGTCATAGAGTGGCATGATTATATTCCTCTGCA 7800 7900
TTGCCAATTTATTCAGCTTTATTTGACTTAGGTGTGCCTTCGTTAGTGACAAATTGCTTTCAAGGAGACAGCCATGCCCCACACTTTGTTGAAAAACAAGTTGCCTTTTGGGATACGGTAAAGCCAGTTGCACTTCAATAATGAATTTCAA 8000
                                                                                   Met Asp Leu His Leu Ile Phe Gly Pro Thr Cys Thr Gly Lys Thr Thr Thr Ala
GGAGACAATATAACCGCCTCTGATAACACAATTCTCTAATATAAAAATCAGTTTGTATTCAATATACTGCAAAAAACTT ATG GAC CTG CAT CTA ATT TTC GGT CCA ACT TGC ACA GGA AAG ACG ACG ACC GCG 8100 8200
Ile Ala Leu Ala Gln Gln Thr Gly Leu Pro Val Leu Ser Leu Asp Arg Val Gln Cys Cys Pro Gln Leu Ser Thr Gly Ser Gly Arg Pro Thr Val Glu Glu Leu Lys Gly Thr
ATA GCT CTT GCC CAG CAG ACA GGG CTT CCA GTC CTT TCG CTT GAT CGG GTC CAA TGC TGT CCT CAA CTA TCA ACC GGA AGC GGA CGA CCA ACA GTG GAA GAA CTG AAA GGA ACG 8300
Thr Arg Leu Tyr Leu Asp Asp Arg Pro Leu Val Glu Gly Ile Ile Ala Ala Lys Gln Ala His His Arg Leu Ile Glu Glu Val Tyr Asn His Glu Ala Asn Gly Gly Leu Ile
ACG CGT CTC TAC CTT GAT GAT CGG CCT CTG GTG GAG GGT ATC ATC GCA GCC AAG CAA GCT CAT CAT AGG CTG ATC GAG GAG GTG TAT AAT CAT GAG GCC AAC GGC GGG CTT ATT 8400
Leu Glu Gly Gly Ser Thr Ser Leu Leu Asn Cys Met Ala Arg Asn Ser Tyr Trp Ser Ala Asp Phe Arg Trp His Ile Ile Arg His Lys Leu Pro Asp Gln Glu Thr Phe Met
CTT GAG GGA GGA TCC ACC TCG TTG CTC AAC TGC ATG GCG CGA AAC AGC TAT TGG AGT GCA GAT TTT CGT TGG CAT ATT ATT CGC CAC AAG TTA CCC GAC CAA GAG ACC TTC ATG 8500
            BamHI
Lys Ala Ala Lys Ala Arg Val Lys Gln Met Leu His Pro Ala Ala Gly His Ser Ile Ile Gln Glu Leu Val Tyr Leu Trp Asn Glu Pro Arg Leu Arg Pro Ile Leu Lys Glu
AAA GCG GCC AAG GCC AGA GTT AAG CAG ATG TTG CAC CCC GCT GCA GGC CAT TCT ATT ATT CAA GAG TTG GTT TAT CTT TGG AAT GAA CCT CGG CTG AGG CCC ATT CTG AAA GAG 8600
Ile Asp Gly Tyr Arg Tyr Ala Met Leu Phe Ala Ser Gln Asn Gln Ile Thr Ala Asp Met Leu Leu Gln Leu Asp Ala Asn Met Glu Gly Lys Leu Ile Asn Gly Ile Ala Gln
ATC GAT GGA TAT CGA TAT GCC ATG TTG TTT GCT AGC CAG AAC CAG ATC ACG GCA GAT ATG CTA TTG CAG CTT GAC GCA AAT ATG GAA GGT AAG TTG ATT AAT GGG ATC GCT CAG 8700
Glu Tyr Phe Ile His Ala Arg Gln Gln Glu Gln Lys Phe Pro Gln Val Asn Ala Ala Ala Phe Asp Gly Phe Glu Gly His Pro Phe Gly Met Tyr
GAG TAT TTC ATC CAT GCG CGC CAA CAG GAA CAG AAA TTC CCC CAA GTT AAC GCA GCC GCT TTC GAC GGA TTC GAA GGT CAT CCG TTC GGA ATG TAT TAG GTTACGCCAGCCCTGCGT 8800 8900
```

Figure 3(ii)

9000
CGCACCTGTCTTCATCTGGATAAGATGTTCGTAATTGTTTTTGGCTTTGTCCCGTTGTGGCAGGGCGGCAAATACTTCCGACAATCCATCGTGTCTTCAAACTTTATGCTGGTGAACAAGTCTTAGTTTCCACGAAAGTATTATGTTAAAT

9100 9200
TTTAAAATTTCGATGTATAATGTGGCTATAATTGTAAAAATAAACTATCGTAAGTGTGCGTGTTATGTATAATTTGTCTAAATGTTTAATATATATCATAGAACGCAATAAATATTAAATATAGCGCTTTTATGAAATATAAATACATCAT

9300
TACAAGTTGTTTATATTTCGGGTACCTTTTCCATTATTTTGCGCAACAAGTCACGGATATTCGTGAAAACGACAAAAACTGCGAAATTTGCGGGCAGTGCCTTCAGTTTTCCTATTAATATTTAGTTTGACACCAGTTGCTATCATTGCG

9400
Met Asp Arg Met Ser Met Ala Arg Gln Gly Trp Leu Val Pro Cys Leu Ser His Gly Lys Asp Asp Gln Leu Gln Gly Glu Leu Ser Glu
GCCAAGCTCAGCTGTTTCTTTTCTTGAAACG ATG GAT CGA ATG AGC ATG GCT CGG CAA GGT TGG CTT GTA CCA TGT CTT TCT CAT GGC AAA GAT GAT CAA CTG CAG GGT GAA CTC TCG GAG

9500
Leu Ser Lys Val Tyr Arg Glu Lys Phe Gln Thr Asp Leu His Thr Lys Ser Gly Asp Ile Ile Asn Pro Gly Gly Glu Phe Leu Tyr Ile Tyr Leu Asp Lys Glu Asn Tyr Arg
CTT TCA AAA GTT TAT CGG GAA AAG TTT CAA ACC GAT CTA CAC ACT AAG TCT GGC GAC ATC ATC AAT CCT GGC GGG GAA TTT TTG TAC ATT TAT CTC GAT AAA GAG AAT TAT CGG

9600 9700
Leu Cys Arg Gln Arg Met Val Leu Val Ser Asn Ala Ser Asp Gly Leu Leu Ala Thr Thr Leu Glu Pro Tyr Ser Asp Gly Tyr Thr Phe Arg Gln Val Arg Ala Gln Leu Gln
TTA TGT CGG CAA AGA ATG GTT CTA GTT TCA AAT GCT TCA GAT GGA TTG CTT GCC ACG ACA CTG GAA CCC TAT TCT GAT GGT TAT ACA TTC CGG CAG GTG AGG GCG CAA CTG CAA

9800
Ala Leu Ser Gly Asp Gly Gly Arg Ile Asn Tyr Ser Lys Asn Glu Tyr Ser Ser Ser Tyr Phe Leu Ala Ile Gln Ala Ser Asn Glu Phe Glu Arg Ile Gly Val Val Arg Asn
GCC TTA AGC GGA GAC GGT GGT CGG ATC AAC TAC AGC AAA AAT GAA TAT TCA TCC TCG TAC TTT CTA GCA ATT CAG GCC AGC AAC GAG TTC GAA CGC ATA GGA GTC GTC CGG AAC

9900
Thr Phe Gly Gln Ser Lys Asp Val Trp Lys Arg Lys Met Pro Ser Ala Gly Gln Pro Leu Asp Tyr Leu Leu Ile Ala Val Gly Cys Ser Ala Phe Leu Pro Glu Ala Ala Leu
ACT TTT GGT CAA AGC AAG GAC GTT TGG AAA AGG AAA ATG CCA AGC GCC GGC CAA CCA TTG GAT TAT CTT TTG ATC GCT GTG GGC TGT TCA GCT TTC CTC CCG GAG GCT GCT CTT

10000
Glu Asp Val Glu Leu Asp Gly Ala Ile
GAA GAC GTG GAG TTA GAT GGT GCT ATC TAG TTTTTATGGGGCGGGATTTCGAAGTCTGTGTCGTTTTGGTTGTTGCATCATAGAGTGCTTATAAAATCTTCTGTTGAATCGAGGAGTGTCTGAAACTTTGTTGAATTAA

10100 10200
TTATTAATGAATAGGACAATTGTGTTCGCTTGTAATTTTCGCCATGTTCATCGTGGGCTGATAAATGTTATATTTAATTCTTCTTCTTGTGTGATCGTGGTGATATTAAAGAGAGTTACAAAATTATTTCGAAACAGGATTTTTCGGCA

10300
ATGATTAGAAATATAAGCTCGTATAGATTATTACCAGGACAGCTTAGAACACTTTTAGAAAAACTAGCGATGGGTGGCGATGTTTGCCGAAAACACAGCCTGCTTTTAGAAGAGGATAACGTTTATTTCGTTACTAAATGACATTGGAA

10400 10500
ACATGCAAAATAACAAAGTCAAGACACACTCAATCACATAGATTAGCCGACTTTATTAGGTGTCGGCGACGGGAA TTA TGC GGA AAG ATC GCA TGA CCC TAA AGC AAT GAT CGG ATA ATT GAT AAG GTT TCC
Ala Ser Leu Asp Cys Ser Gly Leu Ala Ile Ile Pro Tyr Asn Ile Leu Asn Gly

10600
GAT CTG ACA CTC TCT TGG TTC CGC TCC ACC AGC CTC CCC TTC ATC GTC CAT CTC ATC GTC GTC ATC TTC TCC TTC CTC AGT GTA ATT TCC TCC CCG GGC ATA AAA ACC GTA GTT
Ile Gln Cys Glu Arg Pro Glu Ala Gly Gly Ala Glu Gly Glu Asp Asp Met Glu Asp Asp Asp Asp Glu Gly Glu Glu Thr Tyr Asn Gly Gly Arg Ala Tyr Phe Gly Tyr Asn

10700 SmaI
TTC ATT GTT GAT TAT GTC GGT CCC GAA CCG AAC AAA GCA GTT GCT TGG CAT AAT CGC CAA AAA GAA CTG ATT TGT ATG TGG GCC TTG ATA GTA GTT GAC GAG GCC ACC TTG ATC
Glu Asn Asn Ile Ile Asp Thr Gly Phe Arg Val Phe Cys Asn Ser Pro Met Ile Ala Leu Phe Phe Gln Asn Thr His Pro Gly Gln Tyr Tyr Asn Val Leu Gly Gly Gln Asp

10800
GCA CCT TTC CTG CAT AAC TCG ATT CAG GCT GCG TGC ATT CAT CAG CCA CGG CGG TAT TGC AGT TGC CAT TGT TCC GAA GTT AGA GGA GTT GGA CGG TAG GTT TCG ACG CAA GGC
Cys Arg Glu Gln Met Val Arg Asn Leu Ser Arg Ala Asn Met Leu Trp Pro Pro Ile Ala Thr Ala Met Thr Gly Phe Asn Ser Ser Asn Ser Pro Leu Asn Arg Arg Leu Ala

10900
GCA TTG GCG TGC AAT ATC TTC ACG AAG GTA TAC ATA GAC CAG CTC TTG TCG AGA GTG GAT GTA CTC GTC ATC AAA CTC ACC AAG TCT AAT CGC GGA AGG CTG AAA GTA TAC AGT
Cys Gln Arg Ala Ile Asp Glu Arg Leu Tyr Val Tyr Val Leu Glu Gln Arg Ser His Ile Tyr Glu Asp Asp Phe Glu Gly Leu Arg Ile Ala Ser Pro Gln Phe Tyr Val Thr

11000
TTC GGA CAG TAA CGC TCC GAA ATC CGT TCT CGC CTG TTC CAA GCG ACT CTT CAT CTC GCC GGT GCG CAG GAT AAG CGT CAA ATC TCG AAC CTG CCA ATT AGC TAC CGT CAT
Glu Ser Leu Leu Ala Gly Phe Asp Thr Arg Ala Gln Glu Leu Arg Ser Lys Met Glu Gly Thr Arg Leu Ile Leu Thr Leu Asp Arg Val Gln Trp Asn Ala Val Thr Met

11100 11200
CGCAGTGTTGGATGTACTACAAATACCTGCCGCTGGTAAGTCTGAGCCGTTGGTTTTTATATTGACTAAGGAAGCCCATTGACGTCATTGGTGACCGTTTGATGCCGTGTGAACACGACAGATTGATAGCATTGACTAGCGCGTTTG

11300
AATTTTCAGCTGCTGAGCCTCGACATGTTGTCGCAAAATTCGCCCTGGACCCGCCCAACGATTTGTCGTCACTGTCAAGGTTTGACCTGCACTTCATTTGGGGCCCACATACACCAAAAAAATGCTGCATAATTCTCGGGGCAGCAA

11400 11500
GTCGGTTACCCGGCCGCCGTGCTGGACCGGGTTGAATGGTGCCCGTAACTTTCGGTAGAGCGGACGGCCAATACTCAACTTCAAGGAATCTCACCCATGCGCGCCGGCGGGAACCGGAGTTCCCTTCAGTGAGCGTTATTAGTTCGC

11600
CGCTCGGTGTGTCGTAGATACTAGCCCCTGGGGCACTTTTGAAATTTGAATAAGATTTATGTAATCAGTCTTTTAGGTTTGACCGGTTCTGCCGCTTTTTTTAAAATTGGATTTGTAATAATAAAACGCAATTGTTTGTTATTGTGG

11700 11800
CGCTCTATCATAGATGTCGCTATAAACCTATTCAGCACAATATATTGTTTTCATTTTAATATTGTACATATAAGTAGTAGGGTACAATCAGTAAATTGAACGGAGAATATTATTCATAAAAATACGATAGTAACGGGTGATATATT

11900
CATTAGAATGAACCGAAACCGGCGGTAAGGATCTGAGCTACACATGCTCAGGTTTTTTACAACGTGCACAACAGAATTGAAAGCAAATATCATGCGATCATAGGCGTCTCGCATATCTCATTAAAGCAGCTGGAAGATTTGAT TCA AAC
Val

12000
TCC ATT GAG AGC CCT GAC TAT GGC ATT CGC GTT TGA ACC TTC CAG GTT GAG AGA CGA TAG CCC CCT ACC AGT ATG AGA GAG GTC CTC GTT CAG CAC GTC GCT TGC CTC CTG CAC
Gly Asn Leu Ala Arg Val Ile Ala Asn Ala Asn Ser Gly Glu Leu Asn Leu Ser Ser Leu Gly Arg Gly Thr His Ser Leu Asp Glu Asn Leu Val Asp Ser Ala Glu Gln Val

12100
TAC AGA TTT CAT TTC TGG AAC TTG CAC GCC GAT CAC TTC AGC TAT CTC AAC CCA GAG CAC CAA AAT GTG CTT CAC GTC CTC ACT AAG GTA GCG ATG GTT CAT GTT TCT TGG AGT
Val Ser Lys Met Glu Pro Val Gln Val Gly Ile Val Glu Ala Ile Glu Val Trp Leu Val Leu Ile His Lys Val Asp Glu Ser Leu Tyr Arg His Asn Met Asn Arg Pro Thr

12200 12300
TTC GAT ATC AGC GTA GCC CTC AAA GGT CTC ATA GAA TTC TCT TGC GTT GCT GGC GTG CCC ACC ATA CCA TTT TTT GCA GTA TGC GAA ATC CGT CTC GGA CTC AAG TCC AAG CGC
Glu Ile Asp Ala Tyr Gly Glu Phe Thr Glu Tyr Phe Glu Arg Ala Asn Ser Ala His Gly Gly Tyr Trp Lys Lys Cys Tyr Ala Phe Asp Thr Glu Ser Glu Leu Gly Leu Ala

12400
ATT AAC AAT CGT AAG GCG TTC CTC GTC TAT TGC GGT AAC ACG CGT GAT GGC TTG TGG GAC AAA CTT CCG GTA GAA TTT TGG GAT CGG CGA TAT CCC TTG CTC GAT GGT ATC TTT
Asn Val Ile Thr Leu Arg Glu Glu Asp Ile Ala Thr Val Arg Thr Ile Ala Gln Pro Val Phe Lys Arg Tyr Phe Lys Pro Ile Pro Ser Ile Gly Gln Glu Ile Thr Asp Lys

12500
TGC AGC GAG AAT TCC AGC AGG GTG AGC AAC AGG GTT GGT ATT TGA AAA AAA TAT CGA CGC GGG ATT TTG ATA CCA CTG AAG CCG ATT TGG AAA GAG AAT CTC GAA GCC CCC CCT
Ala Ala Leu Ile Gly Ala Pro His Ala Val Pro Asn Thr Asn Ser Phe Phe Ile Ser Ala Pro Asn Gln Tyr Trp Gln Leu Arg Asn Pro Phe Leu Ile Glu Phe Gly Gly Arg

12600
AAC CTC TTC GCT CAA AGC CTG AGT TGA CGC AAC TTC GAA CGT TCT CTT CAC ACT TAG CAT TAG CAC CTG TGC ATT GAC ACG GCG GCA TGC ATA GGG AGA TGT CGT TGC TTC GAT
Val Glu Glu Ser Leu Ala Gln Thr Ser Ala Val Glu Phe Thr Arg Lys Val Ser Leu Met Leu Val Gln Ala Asn Val Arg Arg Cys Ala Tyr Pro Ser Thr Thr Ala Glu Ile

12700
GAC GGC GAT AGG TGC AAA AGC TGG AGT TAA AGT CTG CTT GCA TGC CAG AGA CGT TGC GCT ACC GGG CAA TGC TAC GAG GAC CGA GCT GCT CAG ATT GAA GTT CGC CAA CTC GCA
Val Ala Ile Pro Ala Phe Ala Pro Thr Leu Thr Gln Lys Cys Ala Leu Ser Thr Ala Ser Gly Pro Leu Ala Val Leu Val Ser Ser Ser Leu Asn Phe Asn Ala Leu Glu Cys

12800
AAG AAT TCC TTG CTG GCC CAT GGT CGG GAC CGT AAG AAA AAT GAA CGC CGC GCC TGA AAT CGC TGT TTC AAG ATC ATC CTC CAA TTG CGG CTG AAA GTC GCC TCC ATA GTC CGG
Leu Ile Gly Gln Gln Gly Met Thr Pro Val Thr Leu Phe Ile Phe Ala Ala Gly Ser Ile Ala Thr Glu Leu Asp Asp Glu Leu Gln Pro Gln Phe Asp Gly Gly Tyr Asp Pro

12900
CCC TAC TAG CTC CAA GGA GCC AAG GGA CCT CAC AGA GTT GAA GCT GTT CCT GTT GGA GAT TGG CGC CCA GAT TGA GGA CAC CTG GCC GAG CCT CCG GGC GAG ATC ACC TGC AAG
Gly Val Leu Glu Leu Ser Gly Leu Ser Arg Val Ser Asn Phe Ser Asn Arg Asn Ser Ile Pro Ala Trp Ile Ser Ser Val Gln Gly Leu Arg Arg Ala Leu Asp Gly Ala Leu

13000 13100
AGT AAG AGC CAC GTT TCC CGC ACC CAA AAT TGC CAC TTT AGC CAT TTGGTAGATTGCAAATATAATGGTTTGGCGATTATCCTTGAGGCCACACCTTTAAATAGATCAATGAATGGCAAGTATTGACTTGCGA
Thr Leu Ala Val Asn Gly Ala Gly Leu Ile Ala Val Lys Ala Met

13200
TAATCCTTTAACTTTCTGCCTACCATCAATGTGGATGAGTTGTCGGTAAAAAGGATCCCTGAAAGCGACGTTGGATGTTAACATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCCTTGAGGAA
BamHI

13300 13400
ACTGGTAGCTGTTGTGGGCCTGTGGTCTCAAGATGGATCATTAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGCGATTTGGCCTGTAGACCTCAATTGCGAGCTTTCTAATTTCAAA

13500
CTATTCGGGCCTAACTTTTGGTGTGATGATGCTGACT GGCAGGATATATACCGTTGTAATT TGAGCTCGTGTGAATAGGTCGCTGTGTATGTTTGTTTGATTGTTTCTGTTGGAGTGCAGCCCATTTCACCGGACAAGTCGGCTAGATT

13600
GACTTAGCCCTGATGAACTGCCGAGGGGAAGCCATCTTGAGCTGCGGAATGGGAATGGATTCAGTTG

Figure 3(iii)

Fig. 3. Complete nucleotide sequence of the TL-region of pTiAch5. An uninterrupted sequence of 13 637 bp starting at the *Hind*III site bordering the fragments 14 and 18c covers the whole TL-region. The sequence is displayed in the conventional orientation along with the translation in amino acids for the coding sequences for which experimental evidence exists. The amino acid sequence is above the DNA sequence when transcription occurs from left to right, and below the sequence for the other orientation. The two direct repeats present at both extremities of the TL-DNA are indicated by a closed box. The mRNA start and the polyadenylated sites and signals of transcripts 3, 4 and 7 are indicated by an arrow. The polyadenylation signals of transcripts 3 and 7 are underlined and their polyadenylation sites are indicated by an asterix.

would not produce an easily detected altered phenotype in the transformed plant cells.

Size and position of coding sequences. The sequence between the 24-bp direct repeats was analyzed for possible translational open-reading frames. The 18 largest open-reading frames are presented in Table III. To evaluate which of these open-reading frames are actually used *in vivo*, their position was compared with the known positions of TL-DNA transcripts in octopine crown gall tissues (Willmitzer *et al.*, 1982). Seven of the open-reading frames did correspond with known transcripts. We tested whether or not some of the other open-reading frames might correspond to TL-DNA regions, whose transcripts might have gone undetected, by comparing their position with empty regions in the transcription map. This was the case only for open-reading frame m (Table III). Subsequently, a careful experimental analysis confirmed that this open-reading frame corresponded to an actual transcript (6b) (Willmitzer *et al.*, 1983; Joos *et al.*, 1983). The translation of these eight open-reading frames in amino acids is presented in Figure 3 and their codon usage is listed in Table IV. It was also tested whether open-reading frame p which is derived from the opposite strand of transcript 3 and which might code for a protein of 142 amino acids could correspond to an actual transcript. M13 mp2 phage DNA, containing the small *Eco*RI fragments Ω_1 and Ω_2 (Figure 1) located in the octopine synthase gene, were separately applied on nitrocellulose and hybridized with labeled mRNA isolated from tobacco crown gall tissues. Only the phage DNA spot containing the strand corresponding to transcript 3 (octopine synthase) hybridized with mRNA (data not shown).

We have applied the RNY algorithm described by Shepherd (1981) on the whole sequence of the TL-DNA (data not shown). Eight frames were detected and these correspond to the eight known transcribed regions.

The size and map position of several proteins, expressed by the T-DNA in transformed plant cells, or by the T-region in bacterial cell-free systems, have been recently determined (summarized in Table III). By hybridization selection and translation of T-DNA-encoded mRNA from octopine tumors, three proteins of 39, 27 and 14 kd were detected (Schröder and Schröder, 1982). The largest has been shown to

Table II. DNA sequences homologous to the 24-bp termini sequences

Left terminus sequence	GGCAGGATATATTCAATTGTAAAT	308 bp
	ACCAATTTTTTTTTCAATTCAAAAA	407 bp
	CAGAGTTTATATTCAAAAATCAGT	1024 bp
	CCCAACAGATATACCCTTTGATAT	1293 bp
	CCTTTGATATACTCAATGTATCTT	1307 bp
	CATCTAATCTATTCAGTTTGAAGT	3750 bp
	GGGACAATTAGGTCAATTGTAATA	7777 bp
	TATAATGTGGCTATAATTGTAAAA	9078 bp
	TAAATGTTATATTTAATTCTTCTT	10 131 bp
	CCGGGCATAAAAACCGTAGTTTTC	10 603 bp
	CGGGTGATATATTCATTAGAATGA	11 798 bp
Right terminus sequence	GGCAGGATATATACCGTTGTAATT	13 459 bp

The TL-region sequence was compared with the left and the right terminus sequences using the comparison program written by Schroeder and Blattner (1982). All sequences sharing >50% homology with the terminus sequences were maintained.

Table III. Co-ordinates of open-reading frames of the TL-region DNA

Open region	Nucleotide		First ATG in frame	Σ AA	Mol. wt.		Correspondence
	First	Last			Calculated (d)	Observed (kd)	
a	1054	1740	1060	226	25 635		Transcript 5
b	1569	1135	1512	125	14 310		
c	2726	2307	2687	126	14 219	14	Transcript 7
d	4124	4474	4232	80	8252		
e	4881	3460	4863	467	49 655	49	Transcript 2
f	5155	7476	5209	755	83 815	74	Transcript 1
g	6039	5659	5979	106	12 101		
h	6888	6622	6876	84	10 014		
i	7025	7513	7178	111	12 750		
j	8105	8893	8171	240	26 873	27	Transcript 4
k	8542	8294	8527	77	8858		
l	9344	9970	9395	191	21 335		Transcript 6a
m	11 160	10 453	11 076	207	23 320		Transcript 6b
n	11 142	11 405	11 178	75	8160		
o	11 581	11 092	11 353	86	9375		
p	12 020	12 460	12 032	142	16 455		
q	13 081	11 954	13 030	358	38 665	39	Transcript 3
r	13 203	12 901	13 203	100	11 331		

The table displays all the open-reading frames larger than 75 amino acids. The co-ordinates are those of the first nucleotide following the preceding stop, the last nucleotide of the stop codon and the A of the first ATG in frame. The length of the deduced protein (expressed in amino acids, ΣAA) and its mol. wt. has been calculated and is compared, when possible, with experimental data (Schröder and Schröder, 1982; Schröder *et al.*, 1981, 1983).

Table IV. Codon usage

		Transcripts										Transcripts										Transcripts										Transcripts							
		5	7	2	1	4	6a	6b	3			5	7	2	1	4	6a	6b	3			5	7	2	1	4	6a	6b	3			5	7	2	1	4	6a	6b	3
Phe	UUU	3	5	11	21	2	4	5	8	Ser	UCU	4	1	3	13	1	3	2	6	Tyr	UAU	6	6	8	12	7	7	4	5	Cys	UGU	2	0	3	8	1	3	1	0
	UUC	5	3	6	22	7	3	4	8		UCC	4	2	6	11	1	1	5	5		UAC	2	1	4	11	1	3	6	4		UGC	3	3	4	13	3	0	4	4
Leu	UUA	1	2	9	4	1	3	2	1		UCA	1	3	5	9	1	5	1	5	Stop	UAA	1	1	1	0	0	0	1	0	Stop	UGA	0	0	0	0	0	0	0	1
	UUG	5	3	7	13	6	4	4	7		UCG	4	1	3	6	2	2	0	4		UAG	0	0	0	1	1	1	0	0	Trp	UGG	5	1	2	14	3	2	2	4
	CUU	4	0	7	11	9	6	5	11	Pro	CCU	1	0	7	13	3	1	1	3	His	CAU	2	3	3	13	8	1	1	2	Arg	CGU	1	0	3	6	2	0	2	2
	CUC	5	3	6	16	2	3	1	9		CCC	4	2	8	3	4	1	0	3		CAC	1	1	6	4	2	1	1	3		CGC	4	1	7	6	2	1	4	3
	CUA	2	1	8	5	3	3	1	4		CCA	7	4	11	10	3	3	3	7	Gln	CAA	7	5	5	13	7	8	4	6		CGA	2	0	6	7	3	1	4	0
	CUG	1	5	15	22	6	3	5	4		CCG	2	0	8	12	1	1	4	4		CAG	5	1	3	9	9	3	6	8		CGG	3	1	5	8	3	7	2	3
Ile	AUU	5	2	14	18	8	2	4	8	Thr	ACU	2	4	4	6	1	2	3	6	Asn	AAU	9	4	8	13	4	4	7	8	Ser	AGU	3	1	0	8	1	0	1	2
	AUC	5	2	9	18	7	5	6	9		ACC	1	1	8	8	4	1	2	5		AAC	2	2	12	12	5	3	8	12		AGC	3	2	12	6	3	6	2	7
	AUA	7	2	11	9	1	1	2	5		ACA	5	3	9	14	3	2	2	2	Lys	AAA	8	4	12	17	4	6	0	6	Arg	AGA	1	1	7	5	1	1	3	3
Met	AUG	5	3	5	17	8	5	7	5		ACG	0	0	4	3	5	1	3	7		AAG	6	4	4	17	6	3	1	4		AGG	4	0	1	13	2	2	1	6
Val	GUU	9	1	9	14	3	4	3	8	Ala	GCU	9	1	13	19	6	7	3	10	Asp	GAU	7	2	16	23	6	9	9	6	Gly	GGU	1	2	9	19	4	7	4	6
	GUC	4	1	2	13	2	2	2	6		GCC	1	3	19	14	7	4	2	5		GAC	8	3	12	23	4	4	5	6		GGC	5	2	13	14	2	5	3	8
	GUA	1	2	11	3	0	1	3	3		GCA	3	3	13	18	6	1	5	14	Glu	GAA	7	4	13	23	6	7	9	10		GGA	2	2	13	16	9	3	7	6
	GUG	3	0	8	18	3	3	0	12		GCG	3	1	8	10	4	1	5	10		GAG	1	5	3	15	9	5	8	15		GGG	0	1	6	14	3	1	3	5

There is no general bias in the codon usage of these eight coding sequences taken together, although individually, large deviations do occur. We should note that the transcripts 1, 2, 3, 6a and 6b have a high preference for G as first base (>33.9%) and transcripts 4, 6a, 6b and 7 have a high percentage of A in the second position (>33.2%). No such deviations are noted in the third position.

be octopine synthase (transcript 3). The smallest one was selected with *Hind*III fragment 18 (Figure 1) and corresponds to the translated part of the gene transcript 7. The nucleotide sequences of both transcript 3 and 7 have been described (De Greve *et al.*, 1982a; Dhaese *et al.*, 1983). The third protein (mol. wt. = 27 kd) was observed after hybridization selection both with the partially overlapping fragments *Bam*HI-8 and *Hind*III-1 (Schröder and Schröder, 1982) (Figure 1). The authors suggested that at least part of the coding region is common to both fragments, but we do not find any open-reading frame in this part of the TL-region corresponding to a protein of this size. However, from Table III it appears that the polypeptides encoded by transcript 4 (located in *Hind*III fragment 1; Figure 1) and transcript 5 (located in *Bam*HI fragment 8; Figure 1) have nearly the same mol. wts. (26 873 and 25 635 daltons, respectively). The experimental results obtained by Schröder and Schröder (1982) can be explained if we assume that the observed 27-kd protein bands are in fact different and are encoded by transcripts 4 and 5, respectively.

The TL-region of octopine Ti plasmids expresses four proteins (mol. wt. = 74, 49, 28 and 27 kd) in *Escherichia coli* mini-cells (Schröder *et al.*, 1983). A comparison of the regions expressed in bacteria and the TL-region sequence indicates that three protein-coding regions in the bacteria correspond to three open-reading frames which are transcribed in plants (Table III). The mol. wts. of the polypeptides encoded by transcripts 2 (49 kd) and 4 (27 kd) as calculated from the sequence, are in good agreement with the mol. wts. experimentally observed by Schröder *et al.* (1983) in a bacterial background. However, there is a discrepancy between the calculated (84 kd) and the observed (74 kd) mol. wts. for the protein encoded by transcript 1. Schröder *et al.* (1983) showed that the right-end of the *Bam*HI-8 fragment (Figure 1) in pGV0153 encoded a 66-kd protein, which represents a shortened form of the 74-kd protein. The mol. wt. of this shortened protein calculated from the DNA sequence is 69 kd. Furthermore, deletion of fragment *Hpa*I-14, which is an internal fragment of *Eco*RI fragment 7 (Figure 1) that covers this region, produced a protein of mol. wt. = 53 kd (Schröder *et al.*, 1983). From the DNA sequence we can predict that the first 483 amino acids of transcript 1 will be fused to the last 16 amino acids of transcript 4 in this deletion mutant. The mol. wt. of this fusion protein is 55 kd, in good agreement with the mol. wt. (53 kd) observed by Schröder *et al.* (1983). It is likely, therefore, that the 74-kd protein is indeed encoded by the transcript 1 gene and that the difference in the observed and calculated mol. wts. can be explained by (i) an underestimation of the observed mol. wt. in SDS-polyacrylamide gels, or (ii) proteolytic degradation of this polypeptide in bacteria yielding a shorter protein.

Finally, Schröder *et al.* (1983) observed a 28-kd polypeptide in *E. coli* mini-cells. They located the gene encoding this polypeptide to the left of transcript 4. We do not find an open-reading frame in this region large enough to accommodate this 28-kd protein. Furthermore, no mRNA isolated from crown gall tumors has been observed to hybridize to this region.

Transcription initiation and polyadenylation signals. Comparisons of a multitude of eukaryotic protein-encoding genes have revealed a limited number of consensus sequences potentially involved in RNA polymerase II-mediated transcription. The 'TATA' box or Goldberg-Hogness box (Proudfoot, 1979) is located 25–30 bp upstream from the start site of transcription and is involved *in vivo* in the accurate positioning of the mRNA start site (McKnight and Kingsbury, 1982). The consensus sequence GG(C/T)CAATCT of 'CCAAT' box (Benoist *et al.*, 1980), which appears 40–50 nucleotides upstream of the TATA box, is involved in the regulation of transcription of some eukaryotic genes. By comparing plant genes, a possible regulatory sequence, called AGGA box, was identified by Messing *et al.* (1983). As the transcription of TL-DNA genes is α-amanitin sensitive (Willmitzer *et al.*, 1981) and potential control signals in the 5′ regions of the T-DNA genes (De Greve *et al.*, 1982a; Depicker *et al.*, 1982; Dhaese *et al.*, 1983; Heidekamp *et al.*, 1983), of which the transcription initiation site was accurately determined, have been found resembling those typically used by eukaryotes, we

Table V. Eukaryotic signals present in 5′ and 3′ sequences of the different transcripts

	Position	'CCAAT' box	Position	'TATA' box	Position	Poly(A)$^+$
Consensus sequence		GG^C_TCAATCT		TATA^A_TA^A_T		AATAAA
Transcript 5	909	GGCgAATaT	983	aATAAtA	1912	AATAAT
	935	acgCAATta	1012	TATAAgA	1948	AATAAT
	979	taCCAATaa	1029	TtTATAT		
	1001	GGCCAtTta				
Transcript 7	2800	GtTCAAgCT	2735	TATATAT	2188	AATAAA
Transcript 2	4932	GcgCAAgCT	4909	TATATtT	3281	AATAAT
	4943	caCCAATaa			3297	AATAAT
					3312	AATAAA
					3364	AATAAT
Transcript 1	5092	GcCCAAatT	5175	TATtTAT	7710	AATAAT
	5118	tGTCAAcga			7727	AATAAT
	5144	tcTCAActT				
Transcript 4	8072	ctTCAATaa	8098	aATATAA	9101	AATAAA
	8080	aaTgAATtT	8131	TATAAAA	9169	AATAAA
	8094	aGaCAATaT				
Transcript 6a	9294	GcgaAATtT	9326	TATtAAT	10 030	TATAAA
					10 085	AATGAA
Transcript 6b	11 169	caCCAATga	11 137	TATAAAA	10 260	AATAAT
	11 204	taTCAATCT			10 355	AATAAA
					10 434	AATAAA
Transcript 3	13 114	aCTCAATac	13 088	TATtTAA	11 778	AATAAT
					11 810	AATATA
					11 814	AATGAA

searched for homologies with these putative regulatory sequences in the 5′-untranslated region of the TL-DNA genes. In the 5′-untranslated region of transcript 5, three sequences AATAATA, TATAAGA, and TTTATAT (position 983, 1012 and 1029), sharing homology with the TATA sequence, are located respectively 77, 48 and 31 bp upstream from the translation start codon and are preceded by four 'CCAAT'-like sequences (GGCGAATAT at position 909, ACGCAATTA at 935, TACCAATAA at 979, GGCCATTTA at 1001). Transcript 2 has a TATATTT sequence (position 3460) and two possible CCAAT sequences (GCGCAAGCT at position 4932 and CACCAATAA at 4943). A TATTTAT sequence (position 5175) is located 34 bp upstream from the translation start codon of the gene encoding transcript 1. This TATA box is preceded by three possible CCAAT boxes (positions 5692, 5118, and 5114). The 5′-untranslated region of the gene encoding transcript 6a contains a TATTAAT sequence (position 9326) located 69 bp upstream from the ATG translation codon and a CCAAT sequence (position 9294) located 32 bp upstream from the presumed TATA box. The gene encoding transcript 6b has a TATAAAA sequence (position 11 137) 61 bp upstream from the translation start codon. Two CCAAT sequences (position 11 169 and 11 204) are located upstream of the TATA box at a distance of 32 bp and 67 bp. A summary of the eukaryotic signals found in the 5′-untranslated regions is listed in Table V. However, we did not find sequences in the 5′-untranslated regions of the TL-DNA sharing significant homology with the AGGA box (Messing *et al.*, 1983).

Sequences essential for the *in vivo* expression of eukaryotic genes, however, are located, in most cases, 200–300 bp upstream of the transcription initiation site. From genetic studies, there is evidence that sequences upstream of the TATA and CCAAT boxes are also involved in the *in vivo* expression of the octopine synthase gene (Koncz *et al.*, 1983) in plant cells. We did not find nucleotide sequence homology between this 5′ upstream region of the octopine synthase gene and the 5′ upstream regions of the other TL-DNA genes.

Most eukaryotic protein-encoding transcripts are polyadenylated. The only primary sequence common to the 3′-untranslated region of almost all eukaryotic genes is the hexanucleotide AATAAA (Proudfoot and Brownlee, 1976; Benoist *et al.*, 1980), or a one-base variation of this sequence (Nevins, 1983). This sequence functions in the recognition of the poly(A) addition site (Fitzgerald and Shenk, 1981; Montell *et al.*, 1983). The poly(A) addition sites of the octopine synthase (De Greve *et al.*, 1982a), the nopaline synthase (Depicker *et al.*, 1982), the octopine synthase present in the regenerated plant rGV1 and transcript 7 (Dhaese *et al.*, 1983) are indeed closely preceded by this hexanucleotide signal. In the case of the wild-type octopine synthase and the rGV1 octopine synthase multiple polyadenylation sites have been observed. This was also found to occur in animal genes

(Setzer *et al.*, 1980; Early *et al.*, 1980). We looked for the presence of AATAAA or related sequences in the 3′-untranslated regions of the TL-DNA genes encoding transcripts 5, 2, 1, 6a and 6b. For each gene at least two potential canonical sequences are found. Transcripts 5 and 1 each contain two polyadenylation signals AATAAT (position 1912 and 1948 for transcript 5 and 7710 and 7727 for transcript 1). In transcript 5, these are located at a distance of 172 bp and 208 bp downstream of the stop codon, and those of transcript 1 at 234 bp and 251 bp downstream from the stop codon. The 3′-untranslated region of transcript 2 contains four possible polyadenylation signals: AATAAT (position 3281), AATAAT (3297), AATAAA (3312) and AATAAT (3364), respectively 96, 148, 163 and 180 bp, past the translational stop. In the 3′ region of transcript 6b three polyadenylation signals AATAAT (10 260), AATAAA (10 355), and AATAAA (10 434) are found respectively 193, 98 and 19 bp downstream from the stop codon. Transcript 6a has two sequences: TATAAA (10 030) and AATGAA (10 085) in its 3′ end which are located at a distance of 60 bp and 115 bp downstream from the stop codon. All these data are summarized in Table V.

Translation initiation codons. In eukaryotes, the first AUG of the majority of mRNAs is used as an initiation codon. In the scanning model, two bases (A or G at position −3, G at position +4) flanking the initiation codon (A/GXXAUGG) facilitate the recognition of the functional AUG codon (Kozak, 1981).

Since none of the amino acid sequences of the proteins encoded by the TL-DNA in plant cells have been determined, no experimental data exist concerning the sites used to initiate translation of the plant transcripts. As can be seen in Figure 2, the first AUG following the 'TATA' box is in phase with all the open-reading frames and most likely initiates translation in plants. The first AUG of these plant transcripts are preceded by a very G-poor stretch of DNA and do not contain a Shine-Dalgarno sequence (Shine and Dalgarno, 1974; Stormo *et al.*, 1982). This lack of Gs upstream of eukaryotic initiation codons has already been observed (Kozak, 1981; Sargan *et al.*, 1982).

In the open-reading frames of the genes encoding transcript 5, 7, 2, 4 and 3 the second AUG is located at a distance of 300, 231, 354 and 252 bp, respectively, of the first AUG. In the case of open-reading frames 2 and 4, which are translated in *E. coli* mini-cells (Schröder *et al.*, 1983) these data support the hypothesis that the same translational start is used in bacteria as well as in plant cells. Two AUG codons (positions 11 019 and 11 076) can be used as initiation codon for transcript 6b. Both AUG codons are flanked by a G (position −3) and an A (position +4). Because the initiation codons are equivalent, there is no reason to believe that the first AUG codon is not used as the translational start.

In transcript 6a three AUG codons (position 9395, 9404 and 9410) can be used as initiation codon. The first and the third AUG codons are flanked by two bases which facilitate the recognition of functional AUG codons (Kozak, 1981). Comparison of the TL-DNA sequence of transcript 6a with the corresponding nopaline T-DNA sequence (unpublished data) indicate that in the homologous pTiC58 sequence only the third AUG is conserved. This observation suggests that translation of the octopine transcript 6a starts at the third AUG. However, we cannot exclude that the transcripts 6a encoded by the octopine TL-DNA and the nopaline T-DNA, respectively, have different translational starts.

Transcript 1 also contains three AUG condons in the beginning of the frame (positions 5209, 5260 and 5275). Although we have no data to support that the first AUG is not used as the initiation signal in the plant cells, the possibility exists that the third AUG, which is preceded by a GGTGGA sequence (position 5262) might be preferably used in a bacterial background. The difference in mol. wt. will be 2.3 kd, when calculated from the sequence, and the correspondence with the observed mol. wts. of the shorter polypeptides (53 and 66 kd) (Schröder *et al.*, 1983) and the computed mol. wts. (52.7 and 66.7 kd) are even better.

To solve the question of whether the same translation start codon is used in plant cells and in bacteria, amino acid sequences of both will be needed.

Intervening sequences. A characteristic but not an absolute criterion of eukaryotic genes is the presence of intervening se-

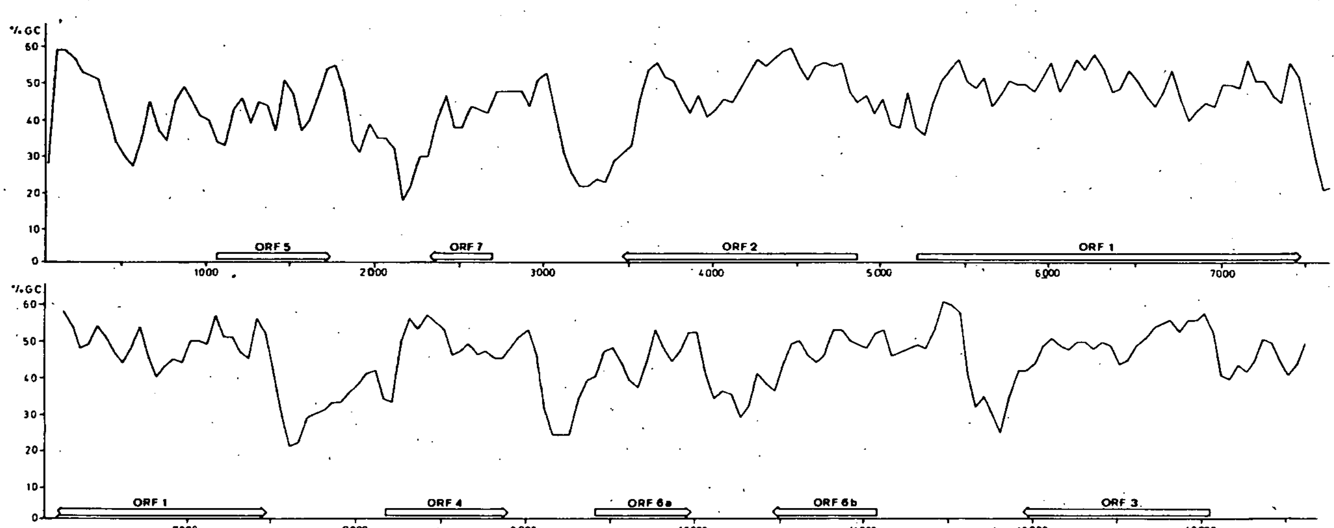


Fig. 4. GC profile of the TL-DNA. A window of 100 bp was slid along the sequence by increments of 50 bp, and its G+C percentage was calculated. The position and size of each known coding region and its orientation is indicated by arrows. The two parts of the figure are contiguous, but the right part of transcript 1 is repeated in the lower figure in order to emphasize the periodicity of the GC content.

quences. To date, several plant nuclear genes have been shown to contain intervening sequences (Sun *et al.*, 1981; Fisher and Goldberg, 1982; Hyldig-Nielsen *et al.*, 1982; Shah *et al.*, 1982), while several others lack intervening sequences (Geraghty *et al.*, 1981; Fisher and Goldberg, 1982; Pedersen *et al.*, 1982). The existence of introns in the coding regions of the different TL-DNA transcripts is very unlikely. Firstly, the open-reading frames correlate well with the sizes of the cytoplasmic polyadenylated transcripts 1, 2, 3, 4, 5, 6a, 6b and 7, determined by Northern analysis (Willmitzer *et al.*, 1982, 1983). Secondly, as discussed above, the sizes of the proteins observed experimentally *in vitro* (Schröder and Schröder, 1982), and in *E. coli* (Schröder *et al.*, 1983) correspond nicely to those calculated from the sequence presented in Figure 3. Furthermore, we have looked without success for sequences fitting with the donor and acceptor consensus sequences proposed by Mount (1982) normally found at the intron-exon junctions.

G+C content. A striking feature of the TL-DNA sequence (Figure 4) is observed when a graphical display of a G+C content profile is plotted. Each functional coding sequence is separated from its neighbours by an AT-rich interval. The 3′-untranslated region of each transcript is very AT-rich, a feature also observed in the 3′-untranslated region of other plant genes, ranging from 24% G+C in the soybean leghemoglobin gene (Hyldig-Nielsen *et al.*, 1982) to 37% G+C in the ribulose-1,5-biphosphate carboxylase gene (Bedbrook *et al.*, 1980). The dip in the G+C profile is less marked between transcripts 1 and 2, possibly because in this case both 5′ ends are very close to one another. Furthermore, these large variations of G+C content can be visualized under the electron microscope by partial denaturation of the Ti plasmid and are limited to the TL-region and the homologous region of the nopaline T-DNA (G. Engler, personal communication).

Conclusions

From the determination and the analysis of the primary structure of the TL-DNA sequence, the following conclusions can be drawn: (i) all the TL-DNA genes contain the signals to be transcribed and translated in plant cells; (ii) the absence of intervening sequences and the compact organization of the genes on the TL-DNA suggest that a maximum amount of genetic information is concentrated in a minimum amount of DNA.

Materials and methods

Enzymes

DNA polymerase I (large fragment, according to Klenow) and T4 polynucleotide kinase were from Boehringer Pharma (Mannheim, FRG).

Restriction enzymes were from Boehringer Pharma (Mannheim, FRG) or New England Biolabs (Beverly, MA, USA), and were used according to the suppliers' instructions.

Bacterial strains and plasmids

Bacterial strains and plasmids are listed in Table I.

Plasmid preparation

Agarose gel electrophoresis, conditions for DNA ligation, and transformation of competent *E. coli* cells were as described by Depicker *et al.* (1980).

Plasmids were prepared from *E. coli* K514 or by CsCl-EtBr equilibrium density gradient centrifugation in cleared SDS lysates (Betlach *et al.*, 1976). The copy number of the pBR derivatives was increased by adding chloramphenicol (170 μg/ml) or spectinomycin (300 μg/ml) to an exponentially growing culture and incubating for a further 15 h.

DNA sequence determination

DNA fragments to be sequenced were labeled at their 5′ ends with [γ-^{32}P]-ATP (>2000 Ci/mmol, Amersham) and T4 polynucleotide kinase (Boehringer, Mannheim, FRG) after treatment with bacterial alkaline phosphatase (Boehringer, Mannheim, FRG); DNA fragments wre labeled at their 3′ ends using either [^{32}P]cordycepin (NEN) and terminal nucleotidyl transferase, or [α-^{32}P]dATP and Klenow polymerase (Boehringer, Mannheim, FRG). The labeled fragments, after secondary restriction, were extracted from low-gelling temperature agarose as described by Wieslander (1979), or, after strand separation, were extracted from acrylamide as described by Maxam and Gilbert (1980).

The five chemical modification and cleavage reactions G, A+G, C+T, C and A+C were performed as described by Maxam and Gilbert (1980). The cleavage products were separated on 8% and 15% gradient acrylamide gels (0.3 mm x 90 cm) containing 8.3 M urea (Sanger and Coulson, 1978). The gels were autoradiographed at −70°C using intensifying screens.

Computer analysis

Routine analysis (restriction sites, overlaps) of the sequencing data was performed on a Cromemco microcomputer using the mapping and comparison programs written by Schroeder and Blattner (1982) for the CP/M operating system. We developed a program along the lines of the RNY algorithm, described by Shepherd (1981) and the programs used to calculate the mol. wt. of the proteins (Table II), the codon usage (Table III), and the GC profile of the sequence (Figure 4). The limited computing ability of our microcomputer did not allow us to perform extensive searches of similarities using the Sellers (1979), or Needleman and Wunsch (1970) algorithms. Imperfect repeats might therefore have escaped. A machine-readable copy of the sequence has been sent for incorporation in the EMBL data base.

Acknowledgements

The authors wish to thank Ms M. De Cock for typing the complete nucleotide sequence and the manuscript, and K. Spruyt and A. Verstraete for photographical work. This research was supported by grants of the 'ASLK-Kankerfonds', 'Fonds voor Geneeskundig Wetenschappelijk Onderzoek' (FGWO 3.0001.82), 'Instituut tot aanmoediging van het Wetenschappelijk Onderzoek in Nijverheid en Landbouw' (IWONL 3849A), and the Services of the Prime Minister (OOA 12052179), and was carried out under Research Contract No. GBI-4-017-B(RS) of the Biomolecular Engineering Programme of the Commission of the European Communities.

References

Bedbrook,J.R., Smith,S.M. and Ellis,R.J. (1980) *Nature*, **287**, 692-697.
Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.*, **8**, 127-142.
Betlach,M.C., Hershfield,V., Chow,L., Brown,W., Goodman,H.M. and Boyer,H.W. (1976) *Fed. Am. Soc. Exp. Biol.*, **35**, 2037-2043.
Bevan,M.W. and Chilton,M.-D. (1982) *J. Mol. Appl. Genet.*, **1**, 539-546.
Caplan,A., Herrera-Estrella,L., Inzé,D., Van Haute,E., Van Montagu,M., Schell,J. and Zambryski,P. (1983) *Science (Wash.)*, **222**, 815-821.
Chilton,M.-D., Drummond,M.H., Merlo,D.J., Sciaky,D., Montoya,A.L., Gordon,M.P. and Nester,E.W. (1977) *Cell*, **11**, 263-271.
Chilton,M.-D., Drummond,M.H., Merlo,D.J. and Sciaky,D. (1978) *Nature*, **275**, 147-149.
Colson,C., Glover,S.W., Symonds,N. and Stacey,K.A. (1965) *Genetics*, **52**, 1043-1050.
De Beuckeleer,M., Lemmers,M., De Vos,G., Willmitzer,L., Van Montagu, M. and Schell,J. (1981) *Mol. Gen. Genet.*, **183**, 283-288.
De Greve,H., Dhaese,P., Seurinck,J., Lemmers,M., Van Montagu,M. and Schell,J. (1982a) *J. Mol. Appl. Genet.*, **1**, 499-512.
De Greve,H., Leemans,J., Hernalsteens,J.P., Thia-Toong,L., De Beuckeleer,M., Willmitzer,L., Otten,L., Van Montagu,M. and Schell,J. (1982b) *Nature*, **300**, 752-755.
Depicker,A., Van Montagu,M. and Schell,J. (1978) *Nature*, **275**, 150-153.
Depicker,A., De Wilde,M., De Vos,G., De Vos,R., Van Montagu,M. and Schell,J. (1980) *Plasmid*, **3**, 193-211.
Depicker,A., Stachel,S., Dhaese,P., Zambryski,P. and Goodman,H.M. (1982) *J. Mol. Appl. Genet.*, **1**, 561-574.
De Vos,G., De Beuckeleer,M., Van Montagu,M. and Schell,J. (1981) *Plasmid*, **6**, 249-253.
Dhaese,P., De Greve,H., Gielen,J., Seurinck,J., Van Montagu,M. and Schell,J. (1983) *EMBO J.*, **2**, 419-426.
Early,P., Rogers,J., Davis,M., Calame,K., Bond,M., Wall,R. and Hood,L. (1980) *Cell*, **20**, 313-319.
Engler,G., Depicker,A., Maenhaut,R., Villarroel-Mandiola,R., Van Montagu,M. and Schell,J. (1981) *J. Mol. Biol.*, **152**, 183-208.
Fisher,R.L. and Goldberg,R.B. (1982) *Cell*, **29**, 651-660.
Fitzgerald,M. and Shenk,T. (1981) *Cell*, **24**, 251-260.
Garfinkel,D.J., Simpson,R.B., Ream,L.W., White,F.F., Gordon,M.P. and

Nester,E.W. (1981) *Cell*, **27**, 143-153.
Geraghty,D., Peifer,M.A., Rubenstein,I. and Messing,J. (1981) *Nucleic Acids Res.*, **9**, 5163-5174.
Heidekamp,F., Dirkse,W.G., Hille,J. and von Ormondt,H. (1983) *Nucleic Acids Res.*, **11**, 6211-6223.
Holsters,M., Villarroel,R., Gielen,J., Seurinck,J., De Greve,H., Van Montagu,M. and Schell,J. (1983) *Mol. Gen. Genet.*, **190**, 35-41.
Hyldig-Nielsen,J.J., Jensen,E.Ø., Paludan,K., Wiborg,O., Garrett,R., Jørgensen,P. and Marcker,K.A. (1982) *Nucleic Acids Res.*, **10**, 689-695.
Joos,H., Inzé,D., Caplan,A., Sormann,M., Van Montagu,M. and Schell,J. (1983) *Cell*, **32**, 1057-1067.
Kahl,G. and Schell,J. (1982) *Molecular Biology of Plant Tumors*, published by Academic Press, NY, 615 pp.
Koncz,C., De Greve,H., André,D., Deboeck,F., Van Montagu,M. and Schell,J. (1983) *EMBO J.*, **2**, 1597-1603.
Kozak,M. (1981) *Nucleic Acids Res.*, **9**, 5233-5252.
Leemans,J., Deblaere,R., Willmitzer,L., De Greve,H., Hernalsteens,J.P., Van Montagu,M. and Schell,J. (1982) *EMBO J.*, **1**, 147-152.
Lemmers,M., De Beuckeleer,M., Holsters,M., Zambryski,P., Depicker,A., Hernalsteens,J.P., Van Montagu,M. and Schell,J. (1980) *J. Mol. Biol.*, **144**, 353-376.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-559.
McKnight,S.L. and Kingsbury,R. (1982) *Science (Wash.)*, **217**, 316-324.
Messing,J., Geraghty,D., Heidecker,G., Hu,N.-T., Kridl,J. and Rubenstein, I. (1983) in Hollaender,A. (ed.), *Genetic Engineering of Plants*, Plenum Press, NY, pp. 211-227.
Montell,C., Fisher,E.F., Caruthers,M.H. and Berle,A.J. (1983) *Nature*, **305**, 600-605.
Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459-472.
Needleman,S.B. and Wunsch,C.D. (1970) *J. Mol. Biol.*, **48**, 443-453.
Nester,E.W. and Kosuge,T. (1981) *Annu. Rev. Microbiol.*, **35**, 531-565.
Nevins,J.R. (1983) *Biochemistry (Wash.)*, **52**, 441-466.
Ohmori,H., Tomizawa,J. and Maxam,A.M. (1978) *Nucleic Acids Res.*, **5**, 1479-1485.
Ooms,G., Bakker,A., Molendijk,L., Wullems,G.J., Gordon,M.P., Nester, E.W. and Schilperoort,R.A. (1982) *Cell*, **30**, 589-597.
Pedersen,K., Devereux,J., Wilson,D.R., Sheldon,E. and Larkins,B.A. (1982) *Cell*, **29**, 1015-1026.
Proudfoot,N.J. (1979) *Nature*, **279**, 376.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211-214.
Sanger,F. and Coulson,A.R. (1978) *FEBS Lett.*, **87**, 107-110.
Sargan,D.R., Gregory,S.P. and Butterworth,P.H.W. (1982) *FEBS Lett.*, **147**, 133-136.
Schroeder,J.L. and Blattner,F.R. (1982) *Nucleic Acids Res.*, **10**, 69-84.
Schröder,G. and Schröder,J. (1982) *Mol. Gen. Genet.*, **185**, 51-55.
Schröder,J., Hillebrand,A., Klipp,W. and Pühler,A. (1981) *Nucleic Acids Res.*, **9**, 5187-5202.
Schröder,G., Klipp,W., Hillebrand,A., Ehring,R., Koncz,C. and Schröder,J. (1983) *EMBO J.*, **2**, 403-409.
Sellers,P.H. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3041.
Setzer,D.R., McGrogan,M., Nunberg,J.H. and Schimke,R.T. (1980) *Cell*, **22**, 361-370.
Shah,D.M., Hightower,R.C. and Meagher,R.D. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1022-1026.
Shepherd,J.C. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1596-1600.
Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342-1346.
Simpson,R.B., O'Hara,P.J., Krook,W., Montoya,A.L., Lichtenstein,C., Gordon,M.P. and Nester,E.W. (1982) *Cell*, **29**, 1005-1014.
Stormo,G.D., Schneider,T.D. and Gold,L.M. (1982) *Nucleic Acids Res.*, **10**, 2971-2996.
Sun,S.M., Slightom,J.L. and Hall,T.C. (1981) *Nature*, **289**, 37-41.
Thomashow,M.F., Nutter,R., Montoya,A.L., Gordon,M.P. and Nester, E.W. (1980) *Cell*, **19**, 729-739.
Wieslander,L. (1979) *Anal. Biochem.*, **98**, 305-309.
Willmitzer,L., Schmalenbach,W. and Schell,J. (1981) *Nucleic Acids Res.*, **9**, 4801-4812.
Willmitzer,L., Simons,G. and Schell,J. (1982) *EMBO J.*, **1**, 139-146.
Willmitzer,L., Dhaese,P., Schreier,P.H., Schmalenbach,W., Van Montagu, M. and Schell,J. (1983) *Cell*, **32**, 1045-1056.
Yadav,N.S., Vanderleyden,J., Bennett,D.R., Barnes,W.M. and Chilton, M.-D. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6322-6326.
Zambryski,P., Depicker,A., Kruger,K. and Goodman,H. (1982) *J. Mol. Appl. Genet.*, **1**, 361-370.
Zambryski,P., Goodman,H., Van Montagu,M. and Schell,J. (1983) in Shapiro,J.A. (ed.), *Mobile Genetic Elements*, Academic Press, NY, pp. 505-535.

Received on 2 January 1984

Plant Molecular Biology

An International Journal on Molecular Biology, Biochemistry and Genetic Engineering

PMB

ISPMB International Society for Plant Molecular Biology

Volume 30, No. 4, February (II) 1996
ISSN 0167-4412 CODEN PMBIDB

Kluwer Academic Publishers

Plant Molecular Biology **30**: 781–793, 1996.

The C-terminal KDEL sequence increases the expression level of a single-chain antibody designed to be targeted to both the cytosol and the secretory pathway in transgenic tobacco

Alexander Schouten[1,*], Jan Roosien[2], Fred A. van Engelen[3], G.A.M. (Ineke) de Jong[1], A.W.M. (Tanja) Borst-Vrenssen[1], Jacoline F. Zilverentant[2], Dirk Bosch[3], Willem J. Stiekema[3], Fred J. Gommers[1], Arjen Schots[2] and Jaap Bakker[1]
[1] *Wageningen Agricultural University, Department of Nematology, P.O. Box 8123, 6700 ES Wageningen, The Netherlands (* author for correspondence);* [2] *Laboratory for Monoclonal Antibodies, P.O. Box 9060, 6700 GW Wageningen, The Netherlands;* [3] *Centre for Plant Breeding and Reproduction Research (CPRO-DLO), Department of Molecular Biology, P.O. Box 16, 6700 AA Wageningen, The Netherlands*

Received 20 July 1995; accepted in revised form 20 November 1995

Key words: Genetic engineering, KDEL retention signal, signal peptide, single-chain antibody, targeting

Abstract

The effects of subcellular localization on single-chain antibody (scFv) expression levels in transgenic tobacco was evaluated using an scFv construct of a model antibody possessing different targeting signals. For translocation into the secretory pathway a secretory signal sequence preceded the scFv gene (scFv-S). For cytosolic expression the scFv antibody gene lacked such a signal sequence (scFv-C). Also, both constructs were provided with the endoplasmic reticulum (ER) retention signal KDEL (scFv-SK and scFv-CK, respectively). The expression of the different scFv constructs in transgenic tobacco plants was controlled by a CaMV 35S promoter with double enhancer. The scFv-S and scFv-SK antibody genes reached expression levels of 0.01% and 1% of the total soluble protein, respectively. Surprisingly, scFv-CK transformants showed considerable expression of up to 0.2% whereas scFv-C transformants did not show any accumulation of the scFv antibody. The differences in protein expression levels could not be explained by the steady-state levels of the mRNAs. Transient expression assays with leaf protoplasts confirmed these expression levels observed in transgenic plants, although the expression level of the scFv-S construct was higher. Furthermore, these assays showed that both the secretory signal and the ER retention signal were recognized in the plant cells. The scFv-CK protein was located intracellularly, presumably in the cytosol. The increase in scFv protein stability in the presence of the KDEL retention signal is discussed.

Introduction

Recent advances in antibody engineering offer various perspectives to endow plants with new properties. Antibodies and antibody fragments can be used to engineer disease resistance, to alter or design metabolic routes with catalytic antibodies, and to study plant growth and develop-

ment by antisense-like approaches [32]. For these applications it is crucial to have functional antibodies located in the proper subcellular compartment. This can be accomplished by providing the antibody with suitable targeting and sorting signals [1].

The engineering of antibodies is facilitated by their domain structure. The domains carrying the antigen-binding loops can be manipulated in different ways to create various biologically active fragments [42]. An interesting and valuable antibody fragment is the single-chain antibody (scFv), in which the variable domains of light and heavy chain are connected by a flexible peptide linker. Through expression of scFvs, several problems inherent to the post-translational processing of complete antibodies, such as assembly of the four subunits, formation of intermolecular disulphide bonds and glycosylation, can be circumvented [15, 17].

Single-chain antibodies have been successfully expressed in plants. Constitutive cytosolic expression of an scFv antibody in tobacco mediated resistance against artichoke mottled crinkle virus [36]. Owen *et al.* [27] and Firek *et al.* [11] reported cytosolic expression and secretion of an anti-phytochrome scFv antibody.

Cytosolic expression of functional scFv antibodies in plants and other eukaryotes [2, 3, 41] is remarkable. The two intramolecular disulphide bridges (one in V_H and one in V_L) which are assumed to be necessary for folding into a stable and functional scFv [14] are expected not to be formed in the reducing environment of the cytosol because of the absence of the enzyme protein disulphide isomerase [13], which catalyses the formation of such bonds.

Despite the reported successes, intracellular expression of scFv antibodies in plants may not be that straightforward. Owen *et al.* [27] reported that only after screening more than 100 transgenic plants, transformed with 'leaderless' scFv constructs, a plant showing an expression level of 0.1% of the total soluble protein fraction was obtained, while transformants expressing the secretory version of the scFv gene produced ten times more scFv protein [11].

The objective of our study was to compare functional expression of scFv proteins in transgenic tobacco plants if targeted to different subcellular compartments. The scFv gene was derived from the heavy and light chain genes of an antibody raised against a cutinase (21C5) of *Botrytis cinerea* [29]. Both with and without signal peptide the expression of this scFv gene greatly improved when the C-terminal endoplasmic reticulum (ER) retention signal peptide, KDEL [28], was added. Possible causes for this strong enhancement of expression and the implications for antibody expression in plants are discussed.

Materials and methods

Bacterial vectors and strains

For cloning of the scFv inserts the bacterial expression vector pHEN1 [19] was modified by substituting the multiple cloning site and deleting the g3p gene (pNEM5). Addition of the KDEL (Lys-Asp-Glu-Leu) coding sequence behind the c-myc tag sequence resulted in pNEM5K. The *Escherichia coli* strains DH5α and TG1 were used for routine cloning and scFv protein expression, respectively.

Plant vectors

The vectors pCPO33, pCPO33T and pCPO35 were used for plant transformations and transient assays. These vectors are closely related to pCPO5 [12] and only differ between the T-DNA borders (Fig. 1). The vector pCPO33 contains a promoter-terminator cassette composed of a truncated cauliflower mosaic virus (CaMV) Cabb B-D 35S promoter (−343/−1) with duplicated enhancer sequence (−343/−90) together with the 38 bp alfalfa mosaic virus (AlMV) RNA4 untranslated leader [33], a polylinker with unique *Nco*I, *Sst*I, *Sma*I and *Bgl*II cloning sites, and the nopaline synthase terminator, respectively. Furthermore, the β-lactamase gene for prokaryotic selection (ampicillin in *E. coli*, carbenicillin in *Agrobacterium tumefaciens*) and the APH(3′)II

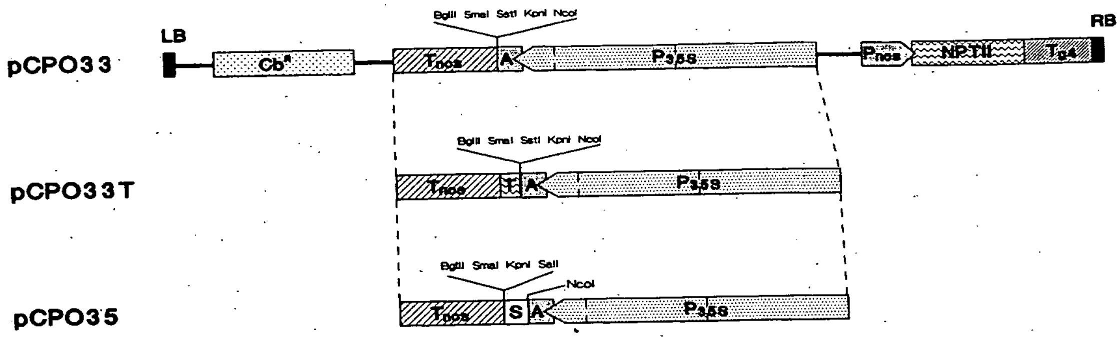


Fig. 1. T-DNA of pCPO33, pCPO33T and pCPO35. LB, left border; RB, right border; A, alfalfa untranslated leader; S, secretory signal sequence; T, c-myc tag; NPTII, neomycin phosphotransferase gene; Pnos, nopaline synthase promoter; Tg4, T-DNA gene 4 terminator; P35S, CaMV 35S promoter with doubled enhancer; Tnos, nopaline synthase terminator; Cb^R, carbenicillin resistance gene.

gene under the control of the nopaline synthase promoter for kanamycin resistance selection at the plant level were located between the T-DNA borders. pCPO33T contains a c-myc tag sequence [25] between the multiple cloning site and nopaline synthase terminator. pCPO35 contains a mouse kappa light chain signal sequence as *Nco*I-*Sal*I fragment between the 35S promoter and the *Kpn*I site. The mouse signal sequence for ER translocation is derived from the kappa light chain, CEA 66E3 [21, 37], and was chosen because minor changes could create a *Sal*I site, which is rare in antibody genes [5, 21]. The signal sequence was made synthetically with an *Nco*I site at the 5′ end (triplet position −24) and a *Sal*I site at the 3′ end (triplet position −3).

Isolation, amplification and cloning of antibody 21C5 variable domains

Isolation of poly(A)$^+$ RNA from 21C5 hybridoma cells [29] was performed by using the QuickPrep *Micro* mRNA purification kit (Pharmacia). First strand cDNA was synthesized using the Pharmacia First Strand cDNA Kit. The variable heavy (V_H) and light domains (V_L) of the 21C5 antibody were amplified through PCR using the following primers: 5′-end primer (H53) 5′-GGTCTCGAGTGTGAGGTCCAGCTGCAACAATCTG-3′ and 3′-end primer (VH33) 5′-ATGCGTTAACCCCGGGTGTTGTTTTGGCTGMRGAGACDGTGAS-3′ for the heavy chain, and 5′-end primer (L5d) 5′-GGTGTCGACGGTGATGTTKTGATGACCCAAA-3′ and 3′-end primer (VK1) 5′-GGCTCGAGTTTGGATTCGGAGCCGGATCCTGAGGATTTACCCTCCCGTTTTATTTCCAGSTTGGTSCCYCC-3′ for the light chain. Primers L5d and H53 contained a *Sal*I and *Xho*I site at their 5′-end, respectively. Primer VH33 was chosen such that, after PCR amplification and digestion with *Sma*I, the V_H domain still contained the initial five triplets of the CH1 domain, encoding Ala-Lys-Thr-Thr-Pro. The primer VK1 carried an *Xho*I site at the 5′-end. Primer VK1 also encodes a sequence for a synthetic linker peptide, adapted with some modifications from Chaudhary *et al.* [5]. For amplification first strand cDNA was denatured at 94 °C for 4 min and subjected to 35 cycles of PCR using Vent DNA Polymerase (New England Biolabs). Each PCR cycle consisted of denaturation at 94 °C for 1 min, annealing at 60 °C for 2 min, and primer extension at 72 °C for 3 min. The amplified fragments were purified from agarose gel, digested with the appropriate restriction enzymes, and ligated simultaneously into *Sal*I/*Sma*I-digested pNEM5 and pNEM5K, resulting in the vectors pNEM-scFv and pNEM-scFv-K, respectively.

The nucleotide sequences of the scFv inserts were verified by the dideoxy chain termination sequencing method [31] on an A.L.F. DNA sequencer (Pharmacia). The sequence encoding the 21C5 scFv was subjected to computer analysis with the Wisconsin GCG software package [9]. From the derived protein sequence the molecular weight was calculated and the algorithm for predicting processing sites for eukaryotic signal sequences was used [39].

Bacterial expression of scFv cassettes

E. coli strain HB 2151 was transformed with pNEM-scFv and pNEM-scFv-K. For the scFv expression assay 5 ml 2 × TY, 1% (w/v) glucose and 100 μg/ml ampicillin was inoculated with a colony containing the appropriate plasmid, and incubated at 30 °C for 16 h. Fresh medium containing 2 × TY, 0.075% (w/v) glucose and 1 μg/ml ampicillin was inoculated with 1/50 volume of the bacterial culture and incubated at 30 °C for 3 h. Then the scFv synthesis was induced by adding isopropyl β-D-thiogalactoside (IPTG) to a final concentration of 1 mM and the incubation was continued for another 4 h. The periplasmic proteins were extracted by osmotic shock [26]. Borate buffer was added to the periplasmic fraction to a final concentration of 0.2 M sodium borate, pH 8.0, and 0.16 M NaCl. The scFvs were purified by affinity chromatography with activated protein A Sepharose (Pharmacia) to which the anti c-myc tag 9E10 monoclonal antibody [25] was covalently attached.

Cloning in plant vectors and tobacco transformation

To generate constructs suitable for cloning in plant vectors without the signal sequence the pNEM-scFv and pNEM-scFv-K vectors were digested with *Nco*I and the ends were filled in with Klenow. After digestion with *Hin*cII at the *Sal*I site the fragments were purified and blunt-end-ligated resulting in the vectors pNEM-scFv-C and pNEM-scFv-CK, respectively. Thus, the *Nco*I site was restored providing the ATG start codon in the proper reading frame. Furthermore, the ATG start codon was placed at position −3 of the mature scFv sequence. The constructs lacking the KDEL sequence were cloned as *Sal*I/*Sma*I (pNEM-scFv) or *Nco*I/*Sma*I (pNEM-scFv-C) fragments into the *Nco*I/*Sma*I digested plant vector pCPO33T. For construction of the scFv-S the *Nco*I/*Sal*I signal sequence fragment was also included in the ligation mixture. The resulting vectors, pCPO-scFv-S and pCPO-scFv-C, had the single chain construct in frame with the c-myc tag sequence. The scFv-K and scFv-CK constructs were cloned as *Sal*I/*Bcl*I (pNEM-scFv-K) and *Nco*I/*Bcl*I (pNEM-scFv-CK) fragments and transferred to the *Sal*I/*Bgl*II-digested pCPO35 and *Nco*I/*Bgl*II-digested pCPO33, respectively. The resulting vectors were designated as pCPO-scFv-SK and pCPO-scFv-CK. All vector-scFv junctions were verified by sequencing.

Tobacco transformation was conducted according to van Engelen *et al.* [37].

Protoplasts

Transient expression assays in tobacco (*N. tabacum* cv. Samsun NN) leaf protoplasts were performed according to the polyethylene glycol procedure as described by Denecke *et al.* [7]. The same protoplasts isolation and culture method was employed to study secretion and retention in transgenic plants. Protoplasts and culture medium were separated by centrifugation and analysed by western blotting experiments and ELISA. For western analysis the proteins present in the culture medium were precipitated with 3 volumes of ethanol. Both protein pellet and protoplasts were dissolved in SDS-PAGE sample buffer (see: protein extraction and analysis). For ELISA the culture medium was diluted 1:1 with PBS, 0.1% Tween, 1% skimmed milk powder and 1 mM 4-(2-aminoethyl)-benzenesulphonyl fluoride hydrochloride (Pefabloc SC, Boehringer) and the protoplasts were lysed in the same buffer. All samples were further treated as described in protein extraction and analysis.

RNA extraction and analysis

For extracting total RNA from plant tissues the guanidine hydrochloride procedure of Logemann *et al.* [23] was used. The RNA concentration was measured spectrophotometrically and northern analysis was carried out according to Sambrook *et al.* [30].

Briefly, 9 μg RNA was separated on a 1.2% (w/v) agarose (Pharmacia) formaldehyde gel. As size marker 1 ng denatured 21C5 scFv DNA and 1 μg of the 0.16–1.77 kb RNA ladder (Life Technologies) were used. After electrophoresis the gel was incubated twice for 15 min in DEPC-treated double distilled water and the RNA was transferred to a Hybond-N + membrane (Amersham) by vacuum blotting, using 20 × SSC as transfer buffer, and cross-linked to the membrane under UV light at 1.5 J/cm^2. The blot was hybridized with [α-^{32}P]dATP-labeled probes at 65 °C for 48 h and further treated as described by Church and Gilbert [6]. The stringency of the final washing was 0.2 × SSC at 65 °C. The blot was first hybridized with a labeled scFv DNA fragment, isolated as *Sal*I-*Sma*I fragment from pNem-scFv. To establish the differences in the amount of total RNA the blot was hybridized with a ribosomal probe. To estimate the molecular sizes the blot was hybridized to labeled cDNA of the RNA ladder. All probes were obtained by random prime labeling [10].

Protein extraction and analysis

Proteins were extracted by grinding tobacco leaves in liquid nitrogen to a fine powder. The powder was transferred to an eppendorf tube and mixed 1:1 (w/v) with SDS-PAGE sample buffer, containing 61 mM Tris-HCl pH 6.8, 2% (w/v) SDS, 12.5% (w/v) glycerol, 1 mM Pefabloc SC (Boehringer). Insoluble plant material was pelleted by centrifugation for 5 min at 13000 rpm. The protein concentration in the supernatant was determined using the BCA method [34]. To the supernatant DTT and bromophenolblue were added to final concentrations of 40 mM and 0.008% (w/v), respectively, and the samples were boiled at 100 °C for 5 min. For non-reducing gel electrophoresis DTT was omitted during sample preparation. Thirty μg of total protein was loaded on a 13% SDS-polyacrylamide gel [22] (BioRad mini protean system). After electrophoresis the proteins were transferred to a PVDF membrane (Millipore) by electroblotting. For immunodetection the membranes were incubated with 1:1000 diluted 9E10 monoclonal antibody, followed by a 1:5000 diluted rat-anti-mouse alkaline phosphatase conjugate (Jackson Immuno Research). Alternatively, a rabbit polyclonal anti-21C5 serum, precleared from antibodies reacting to the constant domains, was used in conjunction with a 1:2500 diluted goat-anti-rabbit alkaline phosphate conjugate (Jackson Immuno Research). The blots were stained in 0.1 M ethanolamine-HCl pH 9.6, supplemented with 4 mM $MgCl_2$, 5-bromo-4-chloro-3-indolyl phosphate (0.06 mg/ml) and nitroblue tetrazolium (0.1 mg/ml). The relative molecular weights of the proteins were estimated with pre-stained low-range molecular weight markers (BioRad).

Purification of native scFv 21C5 antibody from plant extracts was carried out by polytron homogenization of 4 g tobacco leaves in 4 ml 0.2 M sodium borate pH 8.0, containing 0.16 M NaCl and 1 mM Pefabloc SC (Boehringer), in the presence of 200 mg insoluble polyvinylpyrrolidone (Serva). The soluble protein fraction was isolated by centrifugation and filtered through a 0.45 μm filter (Millipore). The scFvs were purified by affinity chromatography with the 9E10 monoclonal antibody coupled to activated protein A Sepharose (Pharmacia) as described previously.

For use in ELISA assays the proteins were extracted by grinding 0.2–0.4 g tobacco leaves in liquid nitrogen to a fine powder. The powder was transferred to an Eppendorf tube and mixed 1:2 (w/v) with PBS-0.1% (v/v) Tween (PBST) and 1 mM Pefabloc SC (Boehringer), and incubated on ice for 5 min. Insoluble material was removed by centrifugation at 13000 × *g*. The supernatant was stored at −80 °C until further use.

The cutinase binding activity of the crude supernatant or purified scFv was determined by

ELISA. A 96-well plate was coated overnight at 4 °C with 1 μg/ml cutinase in 50 mM sodium carbonate, pH 9.6 (100 μl/well). After blocking for 30 min with 200 μl PBST-5% skimmed milk powder per well the plates were washed and 100 μl protein extract per well was added. The plate was incubated for 2 h. To determine the antigen-binding capacity of the scFv antibody preparations, the wells were subsequently washed with PBST, eluted with SDS-PAGE sample buffer, and analysed by immunoblotting under non reducing conditions. Alternatively, for quantitative ELISA, after washing with PBST each well was incubated for another 2 h with 100 μl anti c-myc tag antibody 9E10 (1 ng/μl) in PBST-1% skimmed milk powder. Then, after washing three times with PBST, the wells were incubated for 1 h with alkaline phosphatase conjugated rat-anti-mouse antibody (Jackson Immuno Research), diluted 1:5000 in PBST-1% skimmed milk powder. Finally the wells were washed five times with PBST and 100 μl substrate (0.75 mg/ml p-nitrophenylphosphate in 1 M diethanolamine, pH 9.8) was added and the OD_{405} was monitored. All incubations were carried out at 37 °C.

Results

Construction of the scFv expression cassettes

An scFv gene was constructed containing the variable domains of the 21C5 antibody heavy- (V_H) and light-chain (V_L) genes [37] in the 5′-V_L-linker-V_H-3′ orientation. The end of the V_H region was fused to the *c-myc* tag coding sequence for detection and purification purposes. To enable translocation of the 21C5 scFv to the lumen of the ER it was preceded by a murine κ light-chain signal peptide (scFv-S; Fig. 2A). This sig-

A

```
scFV-S          DVVMTQ==(VL)==IKREGKSSGSGSESKLECEV==(VH)==AAKTTPGRSEQKLISEEDLN
scFV-C       MDGDVVMTQ==(VL)==IKREGKSSGSGSESKLECEV==(VH)==AAKTTPGRSEQKLISEEDLN
scFV-SK         DVVMTQ==(VL)==IKREGKSSGSGSESKLECEV==(VH)==AAKTTPGAAAEQKLISEEDLNDIKDEL
scFV-CK      MDGDVVMTQ==(VL)==IKREGKSSGSGSESKLECEV==(VH)==AAKTTPGAAAEQKLISEEDLNDIKDEL

Ec-scFv     QVDGDVVMTQ==(VL)==IKREGKSSGSGSESKLECEV==(VH)==AAKTTPGAAAEQKLISEEDLNDI
Ec-scFv-K   QVDGDVVMTQ==(VL)==IKREGKSSGSGSESKLECEV==(VH)==AAKTTPGAAAEQKLISEEDLNDIKDEL
```

B

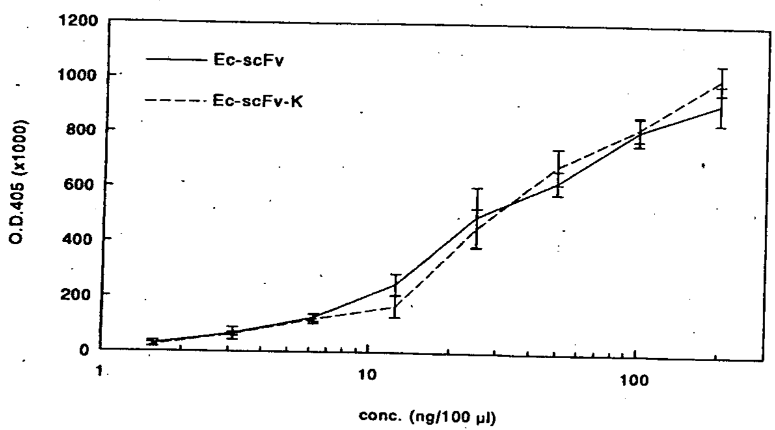


Fig. 2. A. Predicted amino acid sequence of the different mature scFv antibody constructs expressed in plants and bacteria. The V_L and V_H domains are connected by a 16 amino acid linker peptide (bold). The c-myc tag (underlined) is located at the C-terminal end of the peptide. B. Antigen-binding activity of scFv antibodies, isolated from *E. coli*, assayed by ELISA using the anti c-myc antibody. Serial dilutions of purified Ec-scFv and Ec-scFv-K antibodies were incubated in wells coated with 100 ng cutinase. Individual points represent mean values of triplicate trials with standard deviations (error bars).

nal peptide has shown previously to export full-size antibodies efficiently to the plant apoplast [37]. To retain the scFv-S antibody in the ER a C-terminal ER retention signal KDEL was added (scFv-SK; Fig. 2A). In addition two cytosolic versions of the 21C5 scFv (scFv-C and scFv-CK; Fig. 2A) were constructed, which both lacked the ER translocation signal.

To determine if the presence of the KDEL retention signal had any effect on either antigen-binding capacity or detection with the anti c-myc tag antibody, the scFv genes, with and without KDEL sequence, were expressed in *E. coli* (Ec-scFv-K and Ec-scFv, respectively; Fig. 2A). Both scFv genes were preceded by the pelB signal peptide. After affinity purification the Ec-scFv and Ec-scFv-K antibodies showed similar binding properties to the cutinase antigen in an ELISA assay (Fig. 2B). Western blotting followed by immunodetection using the anti c-myc tag antibody revealed proteins of 31 kDa (Fig. 3). Apparently, addition of the KDEL retention signal had no effect on the binding properties of the anti c-myc tag antibody 9E10.

Expression of scFv antibodies in transgenic tobacco leaves

The scFv cassettes were introduced into tobacco by *Agrobacterium* mediated transformation. As a control, transformation also was conducted using the empty vector pCPO33T. Independent kanamycin-resistant transformants were screened by immunoblotting of total protein extracts from leaves. All 43 plants containing the scFv-S constructs showed poor expression. By comparison of the staining intensity on western blot of the plant produced scFv protein with known amounts of bacterially produced scFv, it was estimated that the maximum expression level reached was 0.01% of total protein. No scFv protein was detected in 23 plants containing the scFv-C construct. However, 9 out of 15 scFv-CK transformants showed scFv antibody expression with a maximum level of 0.2% (Fig. 3). Of the 15 scFv-SK transformants 13 were expressing scFv

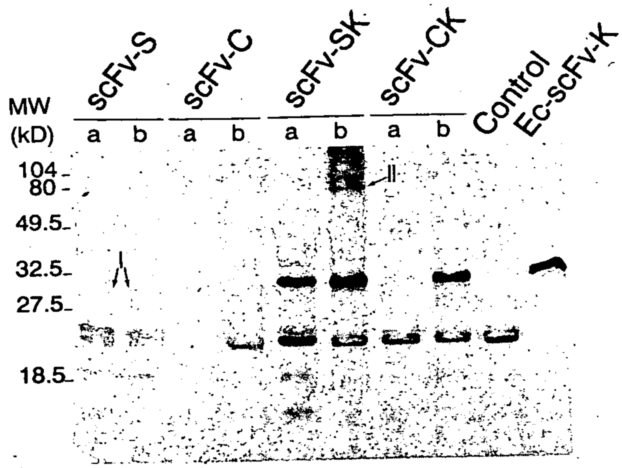


Fig. 3. Western blot analysis of leaf tissue from two independent tobacco transformants (a and b) containing the scFv-S, scFv-C, scFv-SK and scFv-CK cassettes. The lanes marked 'Control' are from a transgenic tobacco plant transformed with the vector pCPO33T. The scFv antibodies were detected using the anti c-myc antibody (9E10). The arrows marked 'I' indicate the scFv-S antibody and the arrow marked 'II' indicates the 65 kDa protein band.

protein, the highest expression level being 1.0% (Fig. 3).

In plants expressing the scFv-SK protein an additional minor product of ca. 65 kDa was detected. To gain more insight into the nature of this 65 kDa band, protein samples were prepared from leaves and analysed under non-reducing conditions. Western blotting showed that under these circumstances the fraction of the 65 kDa protein increased considerably for both scFv-SK and scFv-CK protein preparations (Fig. 4A). This could indicate that in plant cells the cysteine residue present in the linker peptide (Fig. 2A) may have been involved in dimer formation. To determine if both scFv protein and the presumed scFv-dimer had antigen-binding capabilities, purified scFv-CK and scFv-SK antibodies were incubated with immobilized cutinase and analysed after elution under non-reducing conditions (Fig. 4B). Purified bacterially expressed Ec-scFv-K was used as a control. Western blotting of the eluents showed that not only scFv-CK and scFv-SK monomers, but also the 65 kDa proteins bound specifically to the cutinase antigen.

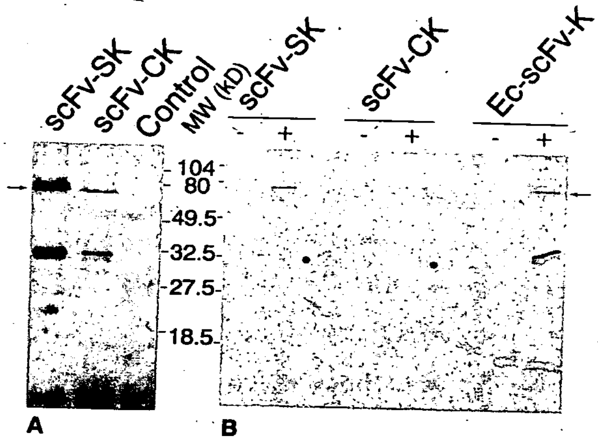


Fig. 4. A. Western blot of proteins from scFv-SK and scFv-CK transgenic tobacco after non-reducing electrophoresis. The scFv antibodies were detected using the anti c-myc antibody (9E10). Arrow indicates the 65 kDa protein band. B. Binding of scFv antibodies to cutinase. Immunoblot of proteins eluted from wells coated with (+) or without (−) cutinase after incubation with scFv-SK and scFv-CK antibodies purified from plants and the Ec-scFv-K antibody purified from *E. coli*. The antibodies were detected with the anti-c-myc antibody (9E10). Asterisk and arrow indicate the scFv antibodies and 65 kDa protein bands, respectively.

Accumulation of scFv mRNA and protein in transgenic tobacco leaves

Since the KDEL retention signal is thought to function only in the secretory pathway the difference in expression level between the scFv-C and scFv-CK was a surprise. Therefore, we investigated whether the differences in protein accumulation between the various constructs could be explained by differences in the steady state mRNA levels. For a number of plants both total RNA and protein was isolated from the same leaf and analysed (Figs. 3 and 5).

Northern blot analysis showed that the scFv transgenic plants accumulated scFv mRNA of the expected size of 1000–1200 bases, albeit in different quantities (Fig. 5A). In addition, a much less abundant mRNA of 1400 bases was detected. The origin of this mRNA is not clear. It was not detected in control plants and therefore may be a read-through product of the scFv messenger

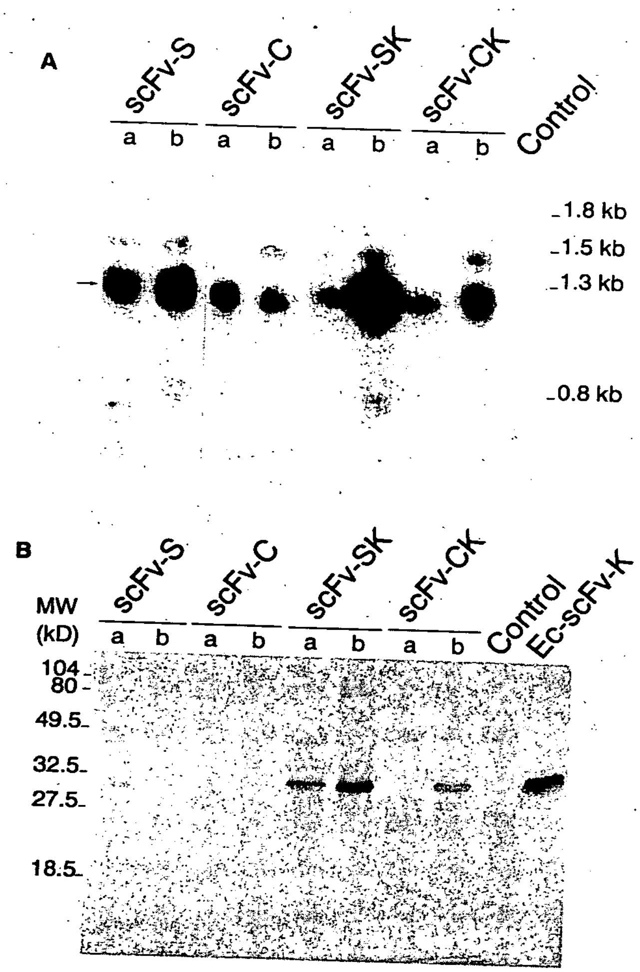


Fig. 5. RNA and protein analysis of leaf tissue from the same independent tobacco transformants as depicted in Fig. 3. A. RNA blot containing total RNA hybridized with a specific scFv probe. Arrow indicates position of scFv transcripts. B. ScFv antibodies detected on western blot using the polyclonal rabbit antiserum against the 21C5 antibody.

RNA. The difference in protein expression level between the different scFv genes could not be explained by various levels of scFv mRNA. The scFv-C mRNA level (Fig. 5A, lane a) was comparable with the scFv-CK mRNA level (Fig. 5A, lane b) but no scFv-C protein was present whereas scFv-CK protein was detected (Figs. 3 and 5B, lanes a and b). Furthermore, a low scFv-SK mRNA level (Fig. 5A, lane a) resulted in a higher scFv protein accumulation than the

relative high scFv-S mRNA level (Fig. 5A, lanes a and b).

To exclude the possibility that the c-myc tag had been removed by plant proteases, thereby affecting our detection procedure, we used both anti-tag antibodies (Fig. 3) and an anti-21C5-Fv rabbit polyclonal antiserum (Fig. 5B) for scFv detection in a number of transgenic plants. Essentially the same results were obtained, indicating that the presence of the complete scFv antibody correlated with the presence of the tag.

The addition of the KDEL retention signal elevated the steady-state levels of the 21C5 scFv antibody, both with and without signal peptide.

Expression of the scFv antibodies in tobacco leaf protoplasts

The four different mature scFv proteins varied slightly in their number of amino acids (Fig. 2A). The calculated sizes of the mature scFv proteins ranged from 30 kDa for the scFv-S to 31 kDa for the scFv-CK. An uncleaved signal peptide would increase the calculated size for the scFv-S and scFv-SK antibodies by 2.5 kDa. On western blot the protein bands showed only minor size differences, the smallest molecule being the scFv-S protein (Fig. 3). This might indicate that the signal peptides of both scFv-S and scFv-SK proteins were recognized and cleaved off during translocation into the ER.

To determine whether the signal peptide and the KDEL retention signal had the predicted effects on scFv protein translocation, transient expression assays were carried out in tobacco protoplasts. Western blot analysis showed differences in the location of the scFv proteins (Fig. 6). As expected, the scFv-S protein was secreted into the incubation medium indicating that the signal peptide was indeed functional. The scFv-SK and scFv-CK proteins were predominantly found inside the protoplasts. The residual presence of scFv-CK and scFv-SK protein in the incubation medium was probably due to cell disruption during the assay, since in a control experiment expressing a β-glucoronidase (GUS) construct

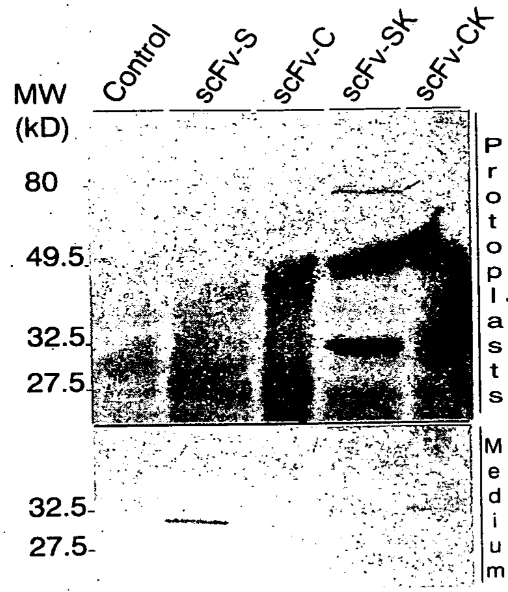


Fig. 6. Western blot analysis of a transient expression assay in tobacco protoplasts transformed with plant vectors containing the scFv-S, scFv-C, scFv-SK and scFv-CK gene cassettes. The 'Control' lane represents the transformation of tobacco protoplasts with the vector pCPO33T. The scFv antibodies present in the cells and incubation medium were detected by the anti c-myc antibody (9E10). Arrow indicates the 65 kDa protein band.

lacking a signal peptide some GUS activity was detected in the medium. From the data obtained with the scFv-SK expression we concluded that the KDEL retention signal was able to retain the scFv-SK antibody inside the protoplasts. This was confirmed by using protoplasts, prepared from the transgenics with a high scFv mRNA level, which showed only intracellular accumulation of the scFv-SK. However, in this case we could not detect the scFv-S antibody, neither in the protoplasts nor in the medium (results not shown).

Discussion

Successful applications for scFv antibodies expressed in plants, including creating resistance against pathogens [36] or altering metabolic pathways (e.g. catalytic antibodies), will to a large

degree depend on the ability to target the scFvs to a particular subcellular compartment and to optimize their expression level. Tavladoraki *et al.* [36] described the successful expression of an scFv antibody directed to the cytosol. Firek *et al.* [11] reported a significant increase in the expression level of an scFv antibody against phytochrome when secreted instead of expressed in the cytosol [27]. They suggested that this difference in expression levels was not the result of a difference in subcellular location but was caused by a destabilizing effect of the phytochrome on the cytosolic scFv [11].

Since no further data on the expression of scFv antibodies in different subcellular compartments of plant cells were available we decided to explore the possibilities of intracellular targeting of scFv antibodies and assess the effect on stability and accumulation. To improve intracellular stability we targeted an scFv antibody away from the cytosol to the potentially more favourable environment of the endoplasmic reticulum (ER) by adding a signal peptide and the tetrapeptide KDEL [8, 16, 40] (scFv-SK). For comparison, a secretory version (scFv-S) of this molecule was used, as well as two cytosolic counterparts, one with and one without the KDEL retention signal (scFv-CK and scFv-C, respectively). The expression level and localization of this scFv-SK antibody were compared with those of the scFv-C, scFv-CK and scFv-S antibodies.

Of the tobacco transformants expressing the scFv-SK cassette, 85% showed a high accumulation of the protein in leaves. In some plants the scFv protein comprised up to 1% of the total soluble protein. Protoplasts prepared from these transgenic plants showed total retention of scFv-SK in the cells. This was confirmed by transient expression assays in tobacco protoplasts. The ScFv-SK antibody was retained intracellularly while a large proportion of the scFv-S antibody was secreted into the culture medium. These results indicated that the signal peptide was functional. Furthermore, they showed that the KDEL retention signal was probably well exposed, recognized by a salvage receptor [35, 38], thereby enabling the scFv antibody to be retained in the ER. When compared with the plants expressing the secreted scFv (scFv-S) the retention in the ER resulted in a 100-fold increase in the amount of detectable scFv antibody. These high accumulation levels cannot be explained by differences in the mRNA levels. It therefore seems that the high level of scFv antibody accumulation is due to its strict localization in the ER and consequently is protected from proteolytic activity further down the secretory pathway, either intra- or extracellularly. Similar results have been obtained with the vacuolar protein vicilin, which also accumulated to a much higher level when retained in the ER [40].

Most striking were the differences in expression levels obtained with the scFv-C and scFv-CK constructs. No transgenic tobacco plants could be found with detectable levels of scFv-C antibody. In contrast to this finding, among the scFv-CK transformants 60% of the plants showed detectable antibody levels. In one plant the scFv-CK protein level reached 0.2% of total soluble protein, which is comparable with previously reported cytosolic expression levels [27, 36]. This difference in expression between the two constructs (scFv-C and scFv-CK) was also found in the transient expression assay. The steady state levels of scFv mRNA indicated that the difference in protein accumulation most likely depended on differences in stability of the protein. This phenomenon is not unique for the anti-cutinase scFv, since we have recently obtained similar results with another scFv antibody (unpublished results).

Presently we can only speculate on the factors which cause these KDEL correlated differences in expression in plants. Assuming that both scFv-C and scFv-CK antibodies are located in the cytosol, it might be possible that the C-terminal extension of the scFv-CK antibody, which is in fact six amino acids long (DIKDEL), protects the scFv from C-terminal degradation by exo-proteinases. This then would suggest that particular exo-proteinases are involved in the breakdown of scFvs. Alternatively, the DIKDEL sequence may indirectly protect the scFv from proteolytic attack via a KDEL mediated interac-

tion with the cytosolic side of the ER salvage receptor [35].

Another explanation for the observed differences could be that expression levels of the scFv are correlated to a different subcellular location. It has been well documented that the expression of normally secreted proteins, particularly those with disulphide bridges, in the cytosol of plant cells is very low [4, 12, 33]. It is therefore not surprising that the scFv-C transformants failed to produce detectable amounts of scFv antibodies. Protein analysis using the algorithm for predicting signal peptidase cleavage sites [39] within the GCG Wisconsin program revealed that both mature scFv-C and scFv-CK proteins did not contain a signal peptide-like sequence in the amino-terminal region. Possibly, the KDEL containing scFvs, even when no signal peptides are added, are directed away from the cytosol to a more favourable location, presumably the ER. The presence of substantial amounts of the 65 kDa protein in the scFv-CK transgenic plants along with its functionality might indicate an ER location. Noteworthy in this respect is the recent suggestion that scFv antibodies targeted to the cytosol of animal cells were actually 'mistranslocated' to the ER [20]. In addition, alternative pathways for secretory proteins, lacking signal peptides, have been put forward [24].

The very low expression from the scFv-S construct in transgenic plants and transgenic protoplasts contrasts with the result obtained in transient expression experiments where we could detect the scFv-S extracellularly. Possibly the protoplasts used for the transient assay were physiologically different from the transgenic scFv-S protoplasts and produced less proteases into the incubation medium. Firek *et al.* [11] reported high expression levels in plants when an anti-phytochrome scFv antibody was being secreted. This difference in stability between different scFv antibodies is not clear but may depend on the amino acid constitution in the variable domains of the scFv antibodies or the linkerpeptide.

Efficient expression of scFv antibodies in different subcellular sites seems feasible. However, it should be kept in mind that successful expression of functional scFvs in the cytosol may only be found under certain conditions, like an scFv amino acid sequence which remains relatively stable [36] or at least can be stabilized by the presence of the antigen [1, 11]. The C-terminal addition of the retention sequence KDEL as a contributing factor for scFv stabilization opens additional opportunities for expressing scFv antibodies in plants.

Acknowledgements

We would like to thank Dr Jesús Salinas for providing the anti-cutinase hybridoma cell line and Jan de Boer for critical reading of the manuscript and helpful discussion. This work was financially supported by grants from the Netherlands Technology Foundation (STW), coordinated through the Foundation for Life Sciences (SLW), and the European Commission (BIO2-CT92-0239).

References

1. Biocca S, Neuberger MS, Cattaneo A: Expression and targeting of intracellular antibodies in mammalian cells. EMBO J 9: 101–108 (1990).
2. Biocca S, Pierandreiamaldi P, Cattaneo A: Intracellular expression of anti-P2L(Ras) single chain Fv fragments inhibits meiotic maturation of xenopus oocytes. Biochem Biophys Res Commun 197: 422–427 (1993).
3. Biocca S, Pierandreiamaldi P, Campioni N, Cattaneo A: Intracellular immunization with cytosolic recombinant antibodies. Bio/technology 12: 396–399 (1994).
4. Bosch D, Smal J, Krebbers E: A trout hormone is expressed, correctly folded and partially glycosylated in the leaves but not the seeds of transgenic plants. Transgenic Res 3: 304–310 (1994).
5. Chaudhary VK, Batra JK, Gallo MG, Willingham MC, FitzGerald DJ, Pastan I: A rapid method of cloning functional variable-region antibody genes in *Escherichia coli* as single-chain immunotoxins. Proc Natl Acad Sci USA 87: 1066–1070 (1990).
6. Church GM, Gilbert W: Genomic sequencing. Proc Natl Acad Sci USA 81: 1991–1995 (1984).
7. Denecke J, Gosselé V, Botterman J, Cornelissen M: Quantitative analysis of transiently expressed genes in plant cells. Meth Mol Cell Biol 1: 19–27 (1989).
8. Denecke J, De Ryke R, Botterman J: Plant and mam-

malian sorting signals for protein retention in the endoplasmic reticulum contain a conserved epitope. EMBO J 11: 2345–2355 (1992).

9. Devereux J, Haerbeli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 12: 387–395 (1984).
10. Feinberg AP, Fogelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6–13 (1983).
11. Firek S, Draper J, Owen MRL, Gandecha A, Cockburn B, Whitelam GC: Secretion of a functional single-chain Fv protein in transgenic tobacco plants and cell suspension cultures. Plant Mol Biol 23: 861–870 (1993).
12. Florack DEA, Dirkse WG, Visser B, Heidekamp F, Stiekema WJ: Expression of biologically active hordothionins in tobacco. Effects of pre and pro-sequences at the amino and carboxyl termini of the hordothionin precursor on mature protein expression and sorting. Plant Mol Biol 24: 83–96 (1994).
13. Freedman RB: Protein disulphide isomerase: multiple roles in the modification of nascent secretory proteins. Cell 57: 1069–1072 (1989).
14. Glockshuber R, Schmidt T, Plückthun A: The disulphide bonds in antibody variable domains: Effects on stability, folding in vitro, and functional expression in *Escherichia coli*. Biochemistry 31: 1270–1279 (1992).
15. Hein MB, Tang Y, McLeod DA, Janda KD, Hiatt A: Evaluation of immunoglobulins from plant cells. Biotechnol Prog 7: 455–461 (1991).
16. Herman EM, Tague BW, Hoffman LM, Kjemtrup SE, Chrispeels MJ: Retention of phytohemagglutinin with carboxyterminal tetrapeptide KDEL in the nuclear envelope and the endoplasmic reticulum. Planta 182: 305–312 (1990).
17. Hiatt A, Cafferkey R, Bowdish K: Production of antibodies in plants. Nature 342: 76–78 (1989).
18. Holliger P, Prospero T, Winter G: 'Diabodies': small bivalent and bispecific antibody fragments. Proc Natl Acad Sci USA 90: 6444–6448 (1993).
19. Hoogenboom HR, Griffiths AD, Johnson KS, Chiswell DJ, Hudson P, Winter G: Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. Nucl Acid Res 19: 4133–4137 (1991).
20. Jiang WR, Venugopal K, Gould EA: Intracellular interference of tick-borne flavivirus infection by using a single-chain antibody fragment delivered by recombinant Sindbis virus. J Virol 69: 1044–1049 (1995).
21. Kabat EA, Wu TT, Reid-Miller M, Perry HM, Gottesman KS: Sequences of proteins of immunological interest. US Department of Health and Human Services, Washington DC (1987).
22. Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685 (1970).
23. Logemann J, Schell J, Willmitzer L: Improved method for the isolation of RNA from plant tissues. Anal Biochem 163: 16–20 (1987).
24. Muesch A, Hartmann E, Rohde K, Rubartelli A, Sitia R, Rapoport TA: A novel pathway for secretory proteins? Trends Biochem Sci 15: 86–88 (1990).
25. Munro S, Pelham HRB: An Hsp-70 like protein in the ER: identity with the 78 kD glucose-regulated protein and immunoglobulin heavy chain binding protein. Cell 46: 291–300 (1986).
26. Neu HC, Heppel LA: The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. J Biol Chem 240: 3685–3692 (1965).
27. Owen M, Gandecha A, Cockburn B, Whitelam G: Synthesis of a functional anti-phytochrome single-chain Fv protein in transgenic tobacco. Bio/technology 10: 790–794 (1992).
28. Pelham HRB: Heat shock and the sorting of luminal ER proteins. EMBO J 8: 3171–3176 (1989).
29. Salinas J: Function of cutinolytic enzymes in the infection of gerbera flowers by *Botrytis cinerea*. Ph.D. thesis, University of Utrecht, The Netherlands (1992).
30. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
31. Sanger F, Nicklen S, Coulson AR: DNA sequencing with the chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
32. Schots A, de Boer J, Schouten A, Roosien J, Zilverentant JF, Pomp H, Bouwman-Smits L, Overmars H, Gommers FJ, Visser B, Stiekema WJ, Bakker J: 'Plantibodies': a flexible approach to design resistance against pathogens. Neth J Plant Path, suppl. 2: 183–191 (1992).
33. Sijmons PC, Dekker BMM, Schrammeijer B, Verwoerd TC, van den Elzen PJM, Hoekema A: Production of correctly processed human serum albumin in transgenic plants. Bio/technology 8: 217–221 (1990).
34. Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ, Klenk DC: Measurement of protein using bicinchoninic acid. Anal Biochem 150: 76–85 (1985).
35. Tang BL, Wong SH, Low SH, Subramaniam VN, Hong WJ: Cytosolic factors block antibody binding to the C-terminal tail of the KDEL receptor. Eur J Cell Biol 65: 298–304 (1994).
36. Tavladoraki P, Benvenuto E, Trinca S, Demartinis D, Cattaneo A, Galeffi P: Transgenic plants expressing a functional single-chain Fv antibody are specifically protected from virus attack. Nature 366: 469–472 (1993).
37. van Engelen FA, Schouten A, Molthoff JW, Roosien, J, Salinas J, Dirkse WG, Schots A, Bakker J, Gommers FJ, Jongsma MA, Bosch D, Stiekema WJ: Coordinate expression of antibody subunit genes yields high levels of functional antibodies in roots of transgenic tobacco. Plant Mol Biol 26: 1701–1710 (1994).
38. Vaux D, Tooze J, Fuller S: Identification of an intercel-

lular receptor for the KDEL retention signal. J Cell Biol 109: 99a (1989).
39. von Heijne G: A new method for predicting signal sequence cleavage sites. Nucl Acids Res 14: 4683–4690 (1986).
40. Wandelt CI, Khan MRI, Craig S, Schroeder HE, Spencer D, Higgins TJV: Vicilin with carboxy-terminal KDEL is retained in the endoplasmic reticulum and accumulates to high levels in the leaves of transgenic plants. Plant J 2: 181–192 (1992).
41. Werge TM, Baldari CT, Telford JL: Intracellular single chain Fv antibody inhibits Ras activity in T-cell antigen receptor stimulated Jurkat cells. FEBS Lett 351: 393–396 (1994).
42. Winter G, Milstein C: Man-made antibodies. Nature 349: 293–299 (1991).

266
25a

YIELDS OF ACPs RETAINED BY THE ANTI-*Escherichia coli* ACP IMMUNOAFFINITY COLUMN[3]

Source of ACP purified by affinity chromatography	Protein (μg)[11]	Units of ACP activity[a]	ACP-dependent fatty acid synthase activity[b]
E. coli	160	203	+
Euglena gracilis strain Z	151	162	+
Euglena gracilis var. *bacillaris*	168	NM[c]	+

[a] One unit of ACP is 1 nmol of $^{14}C_2$ exchanged per 15 min in the malonyl-CoA–CO_2 exchange reaction. The specific activities of the ACPs purified are lower than those reported for freshly reduced, desalted ACP. This is probably due to the absence of a thiol reagent during the chromatographic procedures and to the high salt concentration of the eluents.

[b] ACP is a substrate for the ACP-dependent fatty acid synthase from *Euglena gracilis*. There is no fatty acid biosynthesis in the absence of ACP; therefore, fatty acid biosynthesis is a sensitive indication of the presence of functional ACP.

[c] NM, not measured.

phate, pH 6.2, 0.5 *M* NaCl, until the absorbance (280 nm) of the effluent is zero. Finally, specifically bound ACP is removed by elution with 0.2 *M* glycine, pH 2.8, 0.5 *M* NaCl. After elution of antigen, the immunoadsorbent is equilibrated and stored in 0.01 *M* potassium phosphate, pH 6.2, 0.1 *M* NaCl. It should be washed with several column volumes of the same mixture every 2 weeks.

Properties of the Protein Retained on the Immunoaffinity Column

The elution profiles of *E. coli*, *E. gracilis* strain Z and variety *bacillaris* ACPs from the immunoaffinity column are shown in Fig. 3. In each case, upon application of a crude preparation to the column, protein is specifically retained and later released under acidic conditions. When an excess of pure *E. coli* ACP is processed through the column, a similar result is obtained. The yield achieved in the immunoaffinity chromatography step is a function of the binding capacity of the immunoadsorbent. The yields are identical from one column run to the next (Table) regardless of the excessive amount and stage of purity of the ACP applied to the column. Discontinuous electrophoresis, in 14% acrylamide gels, of the material applied to the affinity column and of the selectively retained protein demonstrates the extent of purification achieved by the single step (Fig. 4).

The material retained from the crude *E. coli* ACP preparation shows a single major band at R_f 0.85 (Fig. 4B), identical to *E. coli* ACP purified according to Majerus *et al.*[4] (Fig. 4C). The single peak retained from E.

gracilis strain Z 70–95% saturation ammonium sulfate fraction exhibits a major band at R_f 0.30, a lesser band at R_f 0.35, and one or two other very faint bands (Fig. 4D). Since the immunoaffinity chromatography is done in the absence of thiol reagents, it is possible that some of the ACP is present as disulfide bridge-linked dimer.[2] The elution pattern of *E. gracilis* var. *bacillaris* ACP from the immunoaffinity column is biphasic (Fig. 3C), but electrophoresis of material in each peak shows a single protein band at R_f 0.39 (Fig. 4G,H). In all cases, upon neutralization and concentration, the retained protein is biologically active in the ACP assays (Table).[11]

The small-scale purifications described here illustrate the potential usefulness of immunoaffinity chromatography in obtaining ACPs from diverse sources. Under the conditions described, stable and reproducible results are obtained through more then 25 runs on a single column.

Acknowledgment

Research support from the following sources is acknowledged: Grants-in-Aid from the American Heart Association; National Science Foundation Grants GB-23905, GB-36016X, and PCM-7904062; and Public Health Service Research Career Development Award 5-KO4-HD-25169 from the National Institute of Child Health and Human Development (M. L. E.-F.).

[1] O. H. Lowry, N. J. Rosebrough, A. L. Farr, and R. J. Randall, *J. Biol. Chem.* **193,** 265 (1951).

[23] Acyl-Acyl Carrier Protein Thioesterase from Safflower

By TOM MCKEON and PAUL K. STUMPF

The acyl-ACP[1] thioesterase catalyzes the hydrolysis of acyl-ACP to free fatty acid and ACP-SH.

$$\text{Acyl-S-ACP} + H_2O \rightarrow \text{acyl-OH} + \text{ACP-SH}$$

The thioesterase is of interest because it terminates the set of biosynthetic reactions that take place on ACP, a water-soluble and lipid-insoluble acyl carrier. Further metabolism of fatty acids appears to occur in membrane systems. Thus, acyl-ACP thioesterase may play an important role in regulating the fatty acid composition of plant tissue.[2]

[1] The abbreviations are ACP, acyl carrier protein; BSA, bovine serum albumin.

[2] W. E. Shine, M. Mancha, and P. K. Stumpf, *Arch. Biochem. Biophys.* **172,** 110 (1976).

Assay Method

Principle. The acyl-ACP thioesterase assay involves the measurement of labeled fatty acid released from labeled acyl-ACP. The free fatty acids are extracted into petroleum ether and counted in a liquid scintillation counter.

Reagents

- Glycine, 0.20 *M*, pH 9.0
- Bovine serum albumin, 10 mg/ml in water
- [^{14}C]Stearoyl-ACP, 10 μM in 0.02 *M* potassium phosphate, pH 6.8 (synthesis described in this volume[3]
- Acetic acid, 1 *M*, in isopropanol with 5 mg/ml each of palmitic and stearic acid
- Petroleum ether, reagent grade, saturated with isopropanol–water, 1 : 1 (v/v)

Procedure. The reaction mixture in a 13 × 100 mm screw-cap tube contains 100 μl of glycine buffer, 70 μl of water, 10 μl of BSA and 10 μl of thioesterase preparation appropriately diluted. The reaction is started by the addition of 10 μl of [^{14}C]stearoyl-ACP, and the reaction is stopped after 10 min at room temperature (20–23°) by the addition of 0.2 ml of the 1 *M* acetic acid reagent. After 10 min, the free fatty acids are extracted with two 2-ml portions of the petroleum ether and the extract is counted.

The assay is linear with respect to time and enzyme concentration up to 40% hydrolysis of substrate.[4] One unit of activity is equal to a rate of hydrolysis of 1 μmol per minute per milligram of protein.

Purification

Acetone Powder Extract. This material is obtained from acetone powder of safflower by the method described for stearoyl-ACP desaturase.[3]

Acid Precipitate. The acetone powder extract is cooled on ice and acidified to pH 5.2 with glacial acetic acid. After 1 hr, the precipitate is centrifuged at 10,000 *g* for 10 min, and the supernatant is adjusted to pH 4.3 with acetic acid. After 1 hr, the precipitate is pelleted and resuspended in one-half the starting volume of 0.02 *M* potassium phosphate buffer, pH 6.8. Insoluble debris is centrifuged out, and the supernatant retains 60% to 80% of the acyl-ACP thioesterase activity (see the table) and less than 5% of the stearoyl-ACP desaturase activity.[4]

ACP-Sepharose 4-B column. This column is run exactly as described

[3] T. McKeon and P. K. Stumpf, this volume [34].

[4] T. McKeon, unpublished data, 1979.

PURIFICATION OF ACYL-ACP THIOESTERASE

Fraction	Total protein[a] (mg)	Total activity (mU)	Specific activity (mU/mg)	Yield (%)	Purification (fold)
Acetone powder extract	400	86	0.22	—	—
Acid precipitate	57	64	1.12	74	5
ACP-Sepharose 4B	.13	23	170	27	770

[a] Protein was determined by the method of O. H. Lowry, N. J. Rosebrough, A. L. Farr, and R. J. Randall, *J. Biol. Chem.* **193,** 265 (1951), using bovine serum albumin as the standard.

for the purification of stearoyl-ACP desaturase. The thioesterase elutes with the 0.30 *M* phosphate wash, with the early fractions containing proportionally more thioesterase and the later fractions more of the desaturase.[4]

Purity. As seen in the table, the acyl-ACP thioesterase is purified 770-fold by this procedure. The stearoyl-ACP desaturase is present as approximately 5% of the bulk protein in the purified preparations of the thioesterase.[4]

Properties

Specificity. Acyl-ACP thioesterase from safflower has a strong preference for oleoyl-ACP as substrate. The preference for substrates under routine assay conditions is oleoyl-ACP > stearoyl-ACP > palmitoyl-ACP with relative rates of 10 : 2 : 1, respectively. The rates of hydrolysis of oleoyl-CoA and stearoyl-CoA are less than 2% of the rate of hydrolysis of the corresponding acyl-ACP.[4]

Stability. Preparations purified through the ACP-Sepharose 4B column step are stable for 3 weeks at 4° when maintained in 1 m*M* DTT.[4]

pH Activity Profile. The thioesterase is half-maximally active at pH 8.5 and pH 10.0 with optimum activity at pH 9.5. The thioesterase has less than 2% maximal activity at pH 6.5 and below, where the stearoyl-ACP desaturase is maximally active.[4]

[redacted]
69

Applied Molecular Genetics of Fungi

Symposium of the British Mycological Society held at the University of Nottingham, April 1990

Edited by
J. F. Peberdy, C. E. Caten, J. E. Ogden & J. W. Bennett

The right of the University of Cambridge to print and sell all manner of books was granted by Henry VIII in 1534. The University has printed and published continuously since 1584.

Published for the British Mycological Society by
CAMBRIDGE UNIVERSITY PRESS
Cambridge
New York Port Chester Melbourne Sydney

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY10011-4211, USA
10 Stamford Road, Oakleigh, Victoria 3166, Australia

First published 1991

Designed and produced by **Page Design,** 48 Alan Road, Stockport SK4 4LE

Printed in Great Britain by the University Press, Cambridge

A catalogue record of this book is available from the British Library

Library of Congress cataloguing in publication data available

ISBN 0-521-41571-3

Contents

Contributors page *vi*

Preface *ix*

1 Gene-transfer systems and vector development for filamentous fungi 1
C. A. M. J. J. van den Hondel & P. J. Punt

2 Strategies for cloning genes from filamentous fungi 29
G. Turner

3 Novel methods of DNA transfer 44
J. W. Watts & N. J. Stacey

4 *Saccharomyces cerevisiae*: a host for the production of foreign proteins 66
J. E. Ogden

5 The molecular biology of *Trichoderma reesei* and its application to biotechnology 85
M. Penttilä, T. T. Teeri, H. Nevalainen & J. K.C. Knowles

6 Expression of heterologous genes in filamentous fungi 103
R. W. Davies

7 Methylotrophic yeasts as gene expression systems 118
R. A. Veale & P. E. Sudbery

8 Strain improvement of brewing yeast 128
E. Hinchliffe

9 Identification of the *Cephalosporium acremonium pbc*AB gene using predictions from an evolutionary hypothesis 145
P. L. Skatrud, J. Hoskins, M. B. Tobin, J. R. Miller, J. S. Wood, S. Kovacevic & S. W. Queener

10 Applications of genetically manipulated yeasts 160
A. W. M. Strasser, Z. A. Janowicz, R. O. Roggenkamp, U. Dahlems, U. Weydemann, A. Merckelbach, G. Gellissen, R. J. Dohmen, M. Piontek, K. Melber & C. P. Hollenberg

11 Molecular biology of fungal plant pathogenicity 170
R. P. Oliver, M. L. Farman, N. J. Talbot & M. T. McHale

Contributors

Joan W. Bennett *Department of Cell & Molecular Biology, 2000 Percival Stern Hall, Tulane University, New Orleans, Louisiana 70118, U.S.A.*

Christopher E. Caten *School of Biological Sciences, University of Birmingham, PO Box 363, Birmingham B15 2TT, U.K.*

Ulrike Dahlems *Rhein Biotech GmbH, Erkratherstrasse 230, 4000 Düsseldorf 1, Germany*

R. Jürgen Dohmen *Institut für Mikrobiologie, Heinrich-Heine-Universität Düsseldorf, Universitätstrasse 1, 4000 Düsseldorf 1, Germany*

R. Wayne Davies *Robertson Institute of Biotechnology, Department of Genetics, University of Glasgow, Glasgow G11 5JS, U.K.*

Mark L. Farman *Norwich Molecular Plant Pathology Group, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, U.K.*

Gerd Gellissen *Rhein Biotech GmbH, Erkratherstrasse 230, 4000 Düsseldorf 1, Germany*

Edward Hinchliffe *Bass Brewers Ltd, 137 High Street, Burton-on-Trent, E. Staffordshire DE14 1JZ, U.K.*

Cornelius P. Hollenberg *Institut für Mikrobiologie, Heinrich-Heine-Universität Düsseldorf, Universitätstrasse 1, 4000 Düsseldorf 1, Germany*

JoAnn Hoskins *Molecular Genetics Department, Lilly Research Laboratories, Lilly Corp. Center, Indianapolis, Indiana 46285, U.S.A.*

Zbigniew A. Janowicz *Rhein Biotech GmbH, Erkratherstrasse 230, 4000 Düsseldorf 1, Germany*

Jonathan K. C. Knowles *VTT Biotechnical Laboratory, POB 202, SF-02151 Espoo, Finland*

Steven Kovacevic *Molecular Genetics Department, Lilly Research Laboratories, Lilly Corp. Center, Indianapolis, Indiana 46285, U.S.A.*

Mark T. McHale *Norwich Molecular Plant Pathology Group, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, U.K.*

Armin Merckelbach *Rhein Biotech GmbH, Erkratherstrasse 230, 4000 Düsseldorf 1, Germany*

James R. Miller *Molecular Genetics Department, Lilly Research Laboratories, Lilly Corp. Center, Indianapolis, Indiana 46285, U.S.A.*

Helena Nevalainen *Research Laboratories, Alko Ltd., POB 350, SF-00101 Helsinki, Finland*

Jill E. Ogden *Delta Biotechnology Ltd., Castle Court, Castle Boulevard, Nottingham NG7 1FD, U.K.*

Richard P. Oliver *Norwich Molecular Plant Pathology Group, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, U.K.*

John F. Peberdy *Department of Botany, School of Biological Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, U.K.*

Merja Penttilä *VTT Biotechnical Laboratory, POB 202, SF-02151 Espoo, Finland*

Michael Piontek *Institut für Mikrobiologie, Heinrich-Heine-Universität Düsseldorf, Universitätstrasse 1, 4000 Düsseldorf 1, Germany*

Peter J. Punt *TNO Medical Biological Laboratory, PO Box 45, 2280 AA Rijswijk, The Netherlands*

Stephen W. Queener *Molecular Genetics Department, Lilly Research Laboratories, Lilly Corp. Center, Indianapolis, Indiana 46285, U.S.A.*

Rainer O. Roggenkamp *Institut für Mikrobiologie, Heinrich-Heine-Universität Düsseldorf, Universitätstrasse 1, 4000 Düsseldorf 1, Germany*

Paul L. Skatrud *Molecular Genetics Department, Lilly Research Laboratories, Lilly Corp. Center, Indianapolis, Indiana 46285, U.S.A.*

Nicholas J. Stacey *John Innes Centre for Plant Science Research, Colney Lane, Norwich NR4 7UH, U.K.*

Alexander W. M. Strasser *Rhein Biotech GmbH, Erkratherstrasse 230, 4000 Düsseldorf 1, Germany*

Peter E. Sudbery *Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, U.K.*

Nicholas J. Talbot *Norwich Molecular Plant Pathology Group, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, U.K.*

Tuula T. Teeri *VTT Biotechnical Laboratory, POB 202, SF-02151 Espoo, Finland*

Matthew B. Tobin *Molecular Genetics Department, Lilly Research*

Geoffrey Turner *Department of Molecular Biology and Biotechnology, The University, Sheffield S10 2TN, U.K.*

Cees A. M. J. J. van den Hondel *TNO Medical Biological Laboratory, PO Box 45, 2280 AA Rijswijk, The Netherlands*

Rosemary A. Veale *ICI Joint Laboratories, Department of Biochemistry, University of Leicester, Leicester LE1 7RH, U.K.*

John W. Watts *John Innes Centre for Plant Science Research, Colney Lane, Norwich NR4 7UH, U.K.*

Ulrike Weydemann *Rhein Biotech GmbH, Erkratherstrasse 230, 4000 Düsseldorf 1, Germany*

John S. Wood *Molecular Genetics Department, Lilly Research Laboratories, Lilly Corp. Center, Indianapolis, Indiana 46285, U.S.A.*

Preface

The interactions of fungi with mankind are both beneficial and harmful and are deeply rooted in the history of human society and agriculture. Over the centuries man has sought to manipulate the growth of fungi to his advantage; the methods used though largely empirical have often been highly successful. Since the initial development of recombinant DNA technology in bacteria in the early 1970s, biology has been undergoing a revolution which is spreading to all organisms, including fungi. This revolution is marked by the emergence of a new discipline, molecular biology, at the interface between biochemistry and genetics. The approach and techniques of molecular biology enable us to ask and answer fundamental questions about many aspects of fungal biology, and open the way to the directed manipulation of fungal metabolism.

This book arises from a symposium on 'Fungal Molecular Biology' held by the British Mycological Society at the University of Nottingham in April 1990. Altogether, there were 29 main papers presented at the symposium, covering a broad range of both fundamental and applied aspects of fungal molecular biology. In considering a book based on the meeting it seemed desirable, given the inevitable restrictions on space and cost, to focus on one or two areas. The editors decided to highlight the rapid development of gene transfer and cloning techniques in fungi and the ways in which these are being exploited in species of economic importance either in biotechnology or as plant pathogens. The 11 contributions in this volume were selected on that basis.

The relevant methodologies for gene manipulations in fungi are described in the first three chapters. In chapter 1 (Van den Hondel & Punt) the development of suitable vectors and gene transfer systems for filamentous fungi discussed and the wide applicability of these techniques to all fungi is clearly established. One point that emerges is that although a basis of classical genetics is useful, it is not essential. A central feature of this new approach to genetic manipulation is the cloning of genes; several strategies are available in filamentous fungi and the most applicable in each situation can be readily identified (Chapter 2, Turner). To date, the technology for introducing vectors into fungal cells has been restricted

uptake into protoplasts. Workers manipulating plant and animal cells have explored more 'dramatic' procedures as described by Watts & Stacey (Chapter 3).

Not surprisingly, progress in yeast molecular biology has been even more rapid than that with filamentous fungi. Several contributions concerning yeast research were included in the symposium to provide a point of reference for possible future developments with the filamentous fungi. Advances with *Saccharomyces cerevisiae* stem, in part, from its importance in brewing, where several opportunities for exploitation of recombinant strains exist (Chapter 8, Hinchliffe), but mainly from previously established fundamental knowledge of biochemistry, cell biology and genetics in this organism. A clear example of building on the latter is the use of *Saccharomyces* as a host for the expression of heterologous proteins (Chapter 4, Ogden). Despite the fact that this fungus secretes only a limited range of proteins naturally, it can be engineered genetically to secrete significant amounts of recombinant proteins. The success with *Saccharomyces* prompted interest in several other yeasts including the methylotrophic species and several systems are now operational (Chapter 7, Veale & Sudbery; Chapter 10, Strasser *et al.*).

Industrially, the filamentous fungi are best known as sources of antibiotics, organic acids and enzymes. Several of the genes encoding biosynthetic enzymes for β-lactam synthesis have been cloned and manipulated; the advances made in this area in *Cephalosporium* (*Acremonium*) are considered by Skatrud *et al.* (Chapter 9). *Trichoderma* species are used commercially as the producers of a range of hydrolytic enzymes which are secreted into the growth medium. The cellulase system has been investigated using molecular genetic techniques and this has led not only to improvements in cellulase production, but also to the exploitation of this fungus as a host for the expression of heterologous proteins (Chapter 5, Penttilä *et al.*). The Aspergilli are of particular importance in fungal molecular biology as they contain both model experimental and industrially important species. Several of these species are the subject of intense study aimed at developing them as hosts for the commercial production of mammalian proteins (Chapter 6, Davies).

The detrimental economic effects of fungi as agents of plant disease are of even greater importance than the beneficial role of fungi in biotechnology. Most phytopathogenic fungi are not amenable to study by the classical methods of genetics and biochemistry and, as a result, the basic mechanism of fungal pathogen-plant host interactions are poorly understood. However, the approach and techniques of molecular genetics bypass many of these difficulties and are transforming knowledge of all aspects of the biology of these fungi. Clearly there is along way to go before we under-

stand the molecular basis of fungal pathogenicity, but sound foundations are being laid as described in the final chapter (Chapter 11, Oliver *et al.*).

The editors of this volume are grateful to the British Mycological Society for providing the means to organise such a timely and interesting symposium and for supporting the publication of this volume. Generous donations towards the costs of the symposium from Bicon Biochemicals, Cambridge University Press, Glaxo Group Research, Pfizer, SmithKline Beecham and Xenova are gratefully acknowledged. We wish to thank all those who contributed to the meeting and, in particular, the authors of the chapters in this volume for their cooperation in preparing the manuscripts for this book in as short a time as possible. Finally, special thanks go to David Moore and Page Design for their help, guidance and great efficiency in producing the book.

J. F. Peberdy
University of Nottingham

C. E. Caten
University of Birmingham

J. E. Ogden
Delta Biotechnology, Nottingham

J. W. Bennett
Tulane University, New Orleans

Chapter 1

Gene transfer systems and vector development for filamentous fungi

Cees A. M. J. J. van den Hondel & Peter J. Punt

Filamentous fungi have a number of properties which make them important both scientifically and economically. The economic importance can be illustrated by the large variety of products that are made by filamentous fungi, such as organic acids (e.g. citric acid), antibiotics (e.g. penicillin and cephalosporin) and numerous industrial enzymes (e.g. glucoamylase). Filamentous fungi are also used as food (mushrooms), food additives (e.g. the meat extender 'Quorn') and condiments (e.g. soy sauce). A severe, negative economic influence of filamentous fungi is their detrimental effect on crop yield. Plant pathogenic fungi cause annual crop losses of billions of pounds. In addition to their economic importance, filamentous fungi have interesting biological properties such as a complex life cycle, cell differentiation, highly regulated metabolic pathways and efficient secretion of proteins which make them attractive as a model for basic biological research of eukaryotic organisms.

In the pre-recombinant DNA period, physiological, biochemical and genetic studies were mainly carried out with *Neurospora crassa* and *Aspergillus nidulans*. Their haploid genomes, rapid life cycles, simple nutrient requirements and well developed genetic systems made them attractive model systems. Hence, it stands to reason that after the introduction of recombinant DNA techniques, systems for molecular genetic analysis were first developed in these intensively studied filamentous fungi. Thereafter, similar molecular techniques have been extended to less amenable species.

A prerequisite for molecular genetic research in filamentous fungi is the availability of a gene transfer system comprising a vector containing a selectable marker and a transformation procedure for introduction of the vector into the fungus. The specific properties of different types of selection markers can be used to design vectors for specific genetic manipulation strategies necessary for molecular genetic studies.

Recently, several excellent reviews have been published about transformation and genetic engineering of filamentous fungi (Fincham, 1989; Timberlake & Marshall, 1989; Goosen, Bos & Van den Broek, 1990).

Table 1.1. Overview of transformation systems used for filamentous fungi

Mycelial treatment	References
Protoplasts	
$CaCl_2$/PEG	Peberdy (1989) and references therein
liposomes	Radford *et al.* (1981)
electroporation	Ward *et al.* (1989); Thomas & Kenerly (1989); Goldman, Van Montagu & Herrera-Estrella (1990)
Intact cells	
Li acetate	Fincham (1989) and references therein; Bej & Perlin (1989)
biolistic	Armaleo *et al.* (1990)

of the gene transfer systems developed. Special attention will be given to some applications of these systems for genetic manipulation in *Aspergillus*.

Gene transfer systems

For genetic manipulation of filamentous fungi a gene transfer system is required that permits introduction of exogenous DNA and selection of those cells that have incorporated this DNA. This selection can be achieved by covalently linking the DNA to a vector which contains a selection marker. Both transformation frequency and type of transformant can be manipulated by using different types of vector.

Transformation procedure

The procedure to obtain DNA-mediated transformed fungal cells comprises the following steps:

- preparation of cells (protoplasts) which are competent to take up (vector) DNA
- treatment of these cells with the DNA
- regeneration of colony forming units
- selection/detection of those cells that have stably incorporated DNA.

A summary of the transformation systems used for filamentous fungi is given in Table 1.1.

Most frequently, protoplasts are used for the introduction of exogenous DNA. These protoplasts are obtained by incubation of mycelium or spores with cell wall-degrading enzymes in the presence of a compound that stabilizes the protoplasts (for an extensive overview of the different procedures and enzymes used, see Peberdy (1989)). Recently, transfor-

mation through electroporation of protoplasts was described (for references, see Table 1.1). Compared to the generally used $CaCl_2$/PEG method, no significant improvement of transformation frequency was observed.

A few reports describe the use of intact cells for transformation. Both incubation of cells with lithium acetate and particle bombardment (chapter 3) have been successfully used for the transformation of filamentous fungi (for references, see Table 1.1). These methods have the obvious advantage that the sometimes laborious protoplast preparation steps can be omitted.

Selection markers

Three types of selectable marker are used for selection of transformed cells: (a) a gene coding for a suppressor tRNA, (b) auxotrophic markers and (c) dominant selectable markers.

To date there is only one example of a suppressor tRNA gene (*su*-8, presumably a mutant tRNA gene) used as selection marker (Brygoo & Debuchy, 1985). Although this type of marker potentially can be used in each fungal strain which contains a suppressible chain termination mutation, no additional reports of the application of suppressor tRNA genes as selection marker have been published.

Auxotrophic markers are the most commonly used method for selection of transformants. Obviously, a prerequisite for their successful use is the presence of the appropriate mutation in the fungus. In Table 1.2 an overview is given of the auxotrophic markers which have been used. As can be seen from this Table, both homologous and heterologous markers can be used for transformation of fungi.

Some of the markers used (e.g. *pyrG*, *niaD* and *trpC*) have proved to be very useful, since they are functional in several species (Table 1.2). Furthermore, both *pyrG* and *niaD* are attractive markers for developing a gene transfer system in genetically poorly characterized fungal species, since the required mutants can be isolated by positive selection. In the case of *pyrG* they can be isolated by resistance against 5-fluoro-orotic acid (Van Hartingsveldt *et al.*, 1987; Goosen *et al.*, 1987) and in the case of *niaD* by resistance against chlorate (Unkles *et al.*, 1989). Since it is possible to select both for and against the mutant and wild-type phenotypes, these markers are also particularly useful for genetic manipulation strategies, such as gene-replacement experiments.

One of the obvious disadvantages of auxotrophic markers is the need to isolate a recipient strain with the appropriate mutation. With dominant selectable markers both wild-type and mutant strains can be transformed. A list of dominant markers which are utilized is given in Table 1.3. Several

Table 1.2. Auxotrophic selectable markers used for homologous and/or heterologous transformation of filamentous fungi.

Marker (species)**	Encoded function	Transformed species*	Reference
acuA$^+$ (*Ustilago maydis*)	acetyl-coA synthase	*Ustilago maydis*	Hargreaves & Turner (1989)
acuD$^+$ (*Aspergillus nidulans*)	isocitrate lyase	*Aspergillus nidulans*	Ballance & Turner (1986)
ade-2$^+$ (*Schizophyllum commune*)	unknown	*Phanerochaete chrysosporium*	Kornegay, Pribnow & Gold (1989)
am$^+$ (*Neurospora crassa*)	glutamate dehydrogenase	*Neurospora crassa*	Kinsey & Rambosek (1984)
amdS$^+$ (*Aspergillus nidulans*)	acetamidase	*Aspergillus nidulans*	Tilburn *et al.* (1983)
argB$^+$ (*Aspergillus nidulans*)	L-ornithine carbamoyl-transferase	*Aspergillus nidulans*	John & Peberdy (1984)
		Aspergillus niger	Buxton, Gwynne & Davies (1985)
inl$^+$ (*Neurospora crassa*)	unknown	*Neurospora crassa*	Akins & Lambowitz (1985)
leu$^+$ (*Mucor circinelloides*)	unknown	*Mucor circinelloides*	Van Heeswijck & Roncero (1984)
met$^+$ (*Aspergillus oryzae*)	unknown	*Aspergillus oryzae*	Iimura *et al.* (1987)
met-2$^+$ (*Ascobolus immersus*)	homoserine-O-trans acetylase	*Ascobolus immersus*	Goyon & Faugeron (1989)
niaD$^+$ (*Aspergillus nidulans*)	nitrate reductase	*Aspergillus niger*	Malardier *et al.* (1989)
		Fusarium oxysporum	Malardier *et al.* (1989)
niaD$^+$ (*Aspergillus niger*)	nitrate reductase	*Aspergillus niger*	Unkles *et al.* (1989)
		Penicillium chrysogenum	Whitehead *et al.* (1989)
niaD$^+$ (*Aspergillus oryzae*)	nitrate reductase	*Aspergillus oryzae*	Unkles *et al.* (1989)
		Aspergillus nidulans	Unkles *et al.* (1989)

Table 1.2. *continued.*

Marker (species)**	Encoded function	Transformed species*	Reference
nic-1$^+$ (*Neurospora crassa*)	unknown	*Neurospora crassa*	Akins & Lambowitz (1985)
pkiA$^+$ (*Aspergillus nidulans*)	pyruvate kinase	*Aspergillus nidulans*	De Graaff, Van den Broek & Visser (1988)
prn$^+$ (*Aspergillus nidulans*)	proline catabolism	*Aspergillus nidulans*	Durrens *et al.* (1986)
pyr-3$^+$ (*Ustilago maydis*)	dihydroorotase	*Ustilago maydis*	Banks & Taylor (1988)
pyr-4$^+$ (*Neurospora crassa*)	orotidine-5'-phosphate decarboxylase	*Aspergillus nidulans*	Ballance, Buxton & Turner (1983)
pyr-6$^+$ (*Ustilago maydis*)	orotidine-5'-phosphate decarboxylase	*Ustilago maydis*	Kronstad *et al.* (1989)
pyrG$^+$ (*Aspergillus nidulans*)	orotidine-5'-phosphate decarboxylase	*Aspergillus nidulans*	Oakley *et al.* (1987)
pyrG/A$^+$ (*Aspergillus niger*)	orotidine-5'-phosphate decarboxylase	*Aspergillus niger*	Van Hartingsveldt *et al.* (1987), Goosen *et al.* (1987)
		Aspergillus nidulans	Van Hartingsveldt *et al.* (1987)
pyrG$^+$ (*Aspergillus oryzae*)	orotidine-5'-phosphate decarboxylase	*Aspergillus oryzae*	De Ruiter-Jacobs *et al.* (1989)
		Aspergillus niger	De Ruiter-Jacobs *et al.* (1989)
pyroA$^+$ (*Aspergillus nidulans*)	unknown	*Aspergillus nidulans*	May *et al.* (1989)
qa-2$^+$ (*Neurospora crassa*)	catabolic dehydroquinase	*Neurospora crassa*	Case *et al.* (1979)
QUTE$^+$ (*Aspergillus nidulans*)	catabolic dehydroquinase	*Aspergillus nidulans*	Da Silva *et al.* (1986)
riboB$^+$ (*Aspergillus nidulans*)	unknown	*Aspergillus nidulans*	Oakley *et al.* (1987)
trp-1$^+$ (*Cochliobolus heterostrophus*)	trifunctional enzyme of tryptophan biosynthesis***	*Aspergillus nidulans*	Turgeon *et al.* (1986)

Table 1.2. *continued.*

Marker (species)**	Encoded function	Transformed species*	Reference
trp-1$^+$ (*Coprinus cinereus*)	tryptophan synthesis	*Coprinus cinereus*	Binninger *et al.* (1987)
trp-1$^+$ (*Schizophyllum commune*)	trifunctional enzyme of tryptophan biosynthesis***	*Schizophyllum commune*	Munoz-Rivas *et al.* (1986)
		Coprinus cinereus	Casselton & De La Fuenta Herce (1989)
trp-1$^+$ (*Neurospora crassa*)	trifunctional enzyme of tryptophan biosynthesis***	*Neurospora crassa*	Kim & Marzluf (1988)
trp-3$^+$ (*Neurospora crassa*)	tryptophan synthetase	*Neurospora crassa*	Vollmer & Yanofsky (1986)
trpC$^+$ (*Aspergillus nidulans*)	trifunctional enzyme of tryptophan biosynthesis***	*Aspergillus nidulans*	Yelton, Hamer & Timberlake (1984)
		Aspergillus niger	Goosen *et al.* (1989)
trpC$^+$ (*Aspergillus niger*)	trifunctional enzyme of tryptophan biosynthesis***	*Aspergillus nidulans*	Horng, Linz & Pestka (1989)
trpC$^+$ (*Phanerochaete chrysosporium*)	trifunctional enzyme of tryptophan biosynthesis***	*Coprinus cinereus*	Casselton & De La Fuente Herce (1989)
trpC$^+$ (*Penicillium chrysogenum*)	trifunctional enzyme of tryptophan biosynthesis***	*Penicillium chrysogenum*	Sánchez *et al.* (1987), Picknett *et al.* (1987)
		Aspergillus nidulans	Picknett *et al.* (1987)
ura-5$^+$ (*Podospora anserina*)	orotidylic acid pyrophosphorylase	*Podospora anserina*	Bégueret *et al.* (1984)

* listed here are the first species that have been transformed with the marker indicated by homologous or heterologous transformation. In several cases other species have subsequently been transformed with the same marker.

** the species from which the marker was isolated is indicated in parentheses.

*** encodes for glutamine amidotransferase, indoleglycerolphosphate synthetase and phosphoribosylanthranilate isomerase.

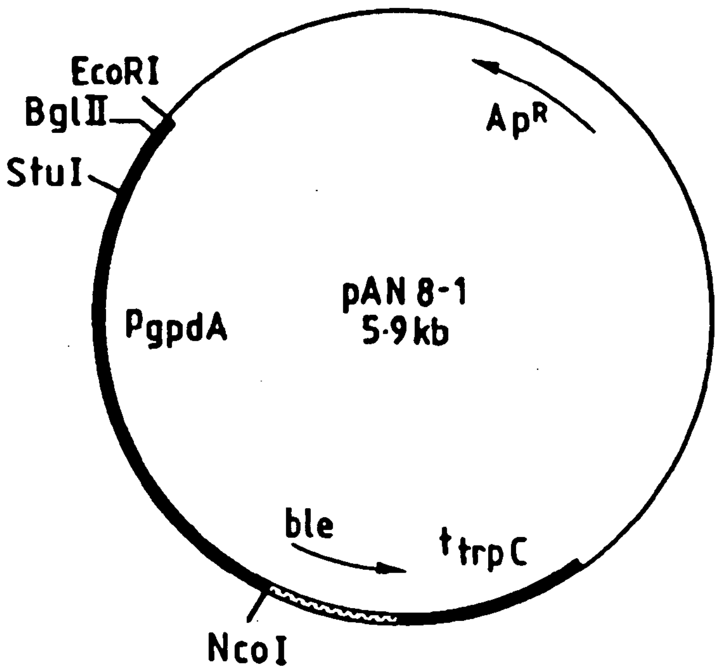


Fig. 1.1. Schematic representation of plasmid pAN8-1, which confers phleomycin resistance after transformation (Mattern, Punt & Van den Hondel, 1988). Thick line represents *A. nidulans* DNA, punctuated line *Streptoalloteichus hindustanus* DNA, and thin line *E. coli* DNA; p_{gpdA}, promoter region of the *gpdA* gene; t_{trpC}, terminator region of the *trpC* gene; *ble*, phleomycin resistance gene; *Ap*^{R}, ampicillin resistance gene. Arrows indicate the direction of transcription.

---

of these markers are 'broad-host range' markers which can be employed in different fungal species. All but one of these markers are based on drug-resistance. They consist of either mutant fungal genes such as benomyl resistant β-tubulin (*benA*, May *et al.*, 1985), or bacterial antibiotic-resistance genes provided with expression signals of filamentous fungi. The only exception is the acetamidase gene of *A. nidulans* (*amdS*, Kelly & Hynes, 1985), which is a nutritional marker. Transformants containing this gene are able to use acetamide or acrylamide as a sole nitrogen and carbon source. In general, fungi cannot readily use these compounds as such.

An example of a vector containing a bacterial resistance gene as selection marker is shown in Fig. 1.1. In this case the *Streptoalloteichus hindustanus* phleomycin resistance (*ble*) gene was introduced in a fungal expression vector containing the promoter region of the highly expressed *A. nidulans gpdA* gene and the terminator region of the *A. nidulans trpC* gene (Punt *et al.*, 1987). This vector and a similar one, pAN7-1, containing the *Escherichia coli* hygromycin B resistance gene have been used for the

Table 1.3. Dominant selectable markers

Marker*	Encoded function	Transformed species	Reference
amdS (*Aspergillus nidulans*)	acetamidase	*Aspergillus niger***	Kelly & Hynes (1985)
bar (*Streptomyces hygroscopicus*)	phosphinothricin acetylase	*Neurospora crassa*	Avalos *et al*. (1989)
benA (*Aspergillus nidulans*)	benomyl resistant β-tubulin	*Aspergillus nidulans*	May *et al*. (1985)
ble (*Escherichia coli*)	phleomycin binding protein	*Penicillium chrysogenum***	Kolar *et al*. (1988)
ble (*Streptoalloteichus hindustanus*)	phleomycin binding protein	*Aspergillus nidulans*/ *Aspergillus niger***	Mattern, Punt & Van den Hondel (1988)
5Fi^r (*Coprinus cinereus*)	5-fluoroindole (feedback) resist-ant anthranilate synthetase	*Coprinus cinereus*	D. M. Burrows, T. J. Elliott & L. A. Casselton (unpublished)
G418^r (*Escherichia coli*)	geneticin/ neomycin/kanamycin phosphotransferase	*Ustilago maydis***	Banks (1983)
hph (*Escherichia coli*)	hygromycin B phosphotransferase	*Cephalosporium acremonium***	Queener *et al*. (1985)
oliC (*Aspergillus nidulans*)	mitochondrial ATP synthase subunit 9	*Aspergillus nidulans*	Ward, Wilkinson & Turner (1986)
oliC (*Aspergillus niger*)	mitochondrial ATP synthase subunit 9	*Aspergillus niger*	Ward *et al*. (1988)
oliC (*Penicillium chrysogenum*)	mitochondrial ATP synthase subunit 9	*Penicillium chrysogenum*	Bull, Smith & Turner (1988)
sul1 (*Escherichia coli*)	dihydropteroate synthetase	*Penicillium chrysogenum*	Carramolino *et al*. (1989)
tub (*Colletotrichum graminicola*)	benomyl resistant β-tubulin	*Colletotrichum graminicola*	Panaccione, McKierman & Hanau (1988)
tub-2 (*Neurospora crassa*)	benomyl resistant β-tubulin	*Neurospora crassa***	Orbach, Porro & Yanofsky (1986)
tubA (*Septoria nodorum*)	benomyl resistant β-tubulin	*Septoria nodorum***	Cooley & Caten (1989)

* the species from which the marker gene was isolated is indicated in parentheses. ** the species listed is the first species transformed with the marker. For the other markers transformation of only one species has been

Types of vector

In general, vectors used for transformation experiments comprise *E. coli* plasmid DNA and the appropriate selectable marker. In most fungal species vector DNA becomes integrated into the genome of the host after transformation. Although considerable effort was undertaken to construct autonomously replicating vectors for *A. nidulans* and *Neurospora crassa*, using a strategy similar to that described for *Saccharomyces cerevisiae* (Stinchcomb, Struhl & Davis, 1979), no autonomous replication of the vector could be detected (Ballance & Turner, 1985; Buxton & Radford, 1984; Paietta & Marzluff, 1985; Van Gorcom, unpublished). In one case, however, a DNA sequence (the *A. nidulans ans1* sequence) which considerably enhances the transformation frequency was isolated. Nevertheless, even this vector did not replicate autonomously (Ballance & Turner, 1985).

For some other species, autonomously replicating vectors were successfully constructed by adding into an integration vector autonomously replicating sequences (ARS) (*Ustilago maydis*, Tsukuda *et al.*, 1988), the chromosomal ends of *Tetrahymena thermophila*, (*Podospora anserina*, Perrot, Barreau & Begueret, 1987), or the termini of naturally occurring linear plasmids of *Nectria haematococca* (*Ustilago maydis*, Samac & Leong, 1989).

In contrast to the results obtained for the ascomycetous fungi, *Neurospora* and *Aspergillus*, in zygomycetous fungi, like *Mucor circinelloides* (van Heeswijck, 1986), *Phycomyces blakesleeanus* (Revuelta & Jayaram, 1986), and *Absidia glauca* (Wostemeyer, Burmester & Weigel, 1987) autonomous replication of vectors was observed in most cases. Autonomous replication was also observed for a filamentous yeast species, *Trichosporon cutaneum* (Glumoff *et al.*, 1989) transformed with pAN7-1 (see above).

Fate of transforming DNA

As already mentioned, in most filamentous fungi vector DNA is integrated into the genome. Biochemical analysis of the DNA of transformants indicates that when a homologous selection marker is used, in general three types of integration events can occur: type I, integration of the vector by homologous recombination; type II, ectopic integration of the vector (or vector sequences) by non-homologous recombination; and type III, gene replacement. For most homologous selectable markers, predominantly homologous interactions (type I and III integrations) occur. However, in some cases type II transformants are preferentially found, e.g. in *A. nidulans* with the *amdS* gene (Wernars *et al.*, 1985) or the *prn* gene cluster (Durrens *et al.*, 1986), and in *Ascobolus immersus* with

Table 1.4. Fungal species successfully transformed with the vectors pAN7-1 and/or pAN8-1

Transformed species	Vector		Reference
	pAN7-1	pAN8-1	
Acremonium chrysogenum	+	ND	A. W. Smith, M. Ramsden & J. F. Peberdy (unpubl.)
Aspergillus nidulans	+	+	Punt *et al.* (1987)
Aspergillus niger	+	+	Punt *et al.* (1987)
Aspergillus ficuum	+	ND	Mullaney, Punt & Van den Hondel (1988)
Aspergillus oryzae	–	+	Mattern, Punt & Van den Hondel (1988)
Aspergillus giganteus	+	ND	Wnendt, Jacobs & Stahl (1990)
Claviceps purpurea	+	ND	Comino *et al.* (1989)
Cryphonectria parasitica	+	ND	Churchill *et al.* (1990)
Curvularia lunata	+	ND	Osiewacz & Weber (1989)
Fulvia fulvum	+	+	Oliver *et al.* (1987)
Fusarium culmorum	+	ND	H. Curragh, R. Marchant, H. Mooibroek & J. G. H. Wessels (unpubl.)
Leptosphaeria maculans	+	ND	Farman & Oliver (1988)

predominantly type II transformants are observed when the *TRP-1* marker is used (Binninger *et al.*, 1987). Transformation of *Ascobolus immersus* with vector DNA linearised by cutting within the marker sequence or with circular single-stranded vector DNA preferentially results in type I integration events (Goyon & Faugeron, 1989).

In the case of heterologous selectable markers integration will always occur through non-homologous recombination, seemingly at random sites in the genome.

Genetic manipulation

The availability of different gene transfer systems with different charac-

Table 1.4. *continued.*

Transformed species	Vector pAN7-1	pAN8-1	Reference
Neurospora crassa	+	ND	Staben *et al.* (1989)
Penicillium chrysogenum	–	+	Kolar *et al.* (1988)
Penicillium roquefortii	+	ND	N. Durand, P. Reymond & M. Fevre (unpubl.)
Pseudocercosporella herpotrichoides	+	ND	Blakemore *et al.* (1989)
Schizophyllum commune	+	ND	Mooibroek *et al.* (1990)
Septoria nodorum	+	ND	Cooley *et al.* (1988)
Talaromyces emersonii	ND	+	S. Jain, H. Durand & G. Tiraby (unpubl.)
Trichoderma harzianum	+	ND	Goldman, Van Montagu & Herrera-Estrella (1990); C. J. Ulhoa, M. H. Vainstein & J. F. Peberdy (unpubl.)
Trichoderma hamatum	+	ND	C. J. Ulhoa, M. H. Vainstein & J. F. Peberdy (unpubl.)
Trichoderma viride	+	ND	Herrera-Estrella, Goldman & Van Montagu (1990)
Trichosporon cutaneum	+	+	Glumoff *et al.* (1989)

interesting processes by isolation, characterisation and functional analysis of the genes and gene products involved. To perform these studies, specific vectors are constructed which facilitate genetic manipulation such as cloning of a gene by complementation of a mutation, gene disruption or gene replacement, and analysis of expression signals *in vivo*.

To illustrate the possibilities of genetic manipulation for molecular genetic studies examples will be given of research on *Aspergillus* that is in progress in our laboratory. The first example concerns experiments that have been performed to prove that a gene encoding a functional benzoate-*p*-hydroxylase gene of *Aspergillus niger* was cloned. In the second example, the application of different expression analysis vectors for a functional

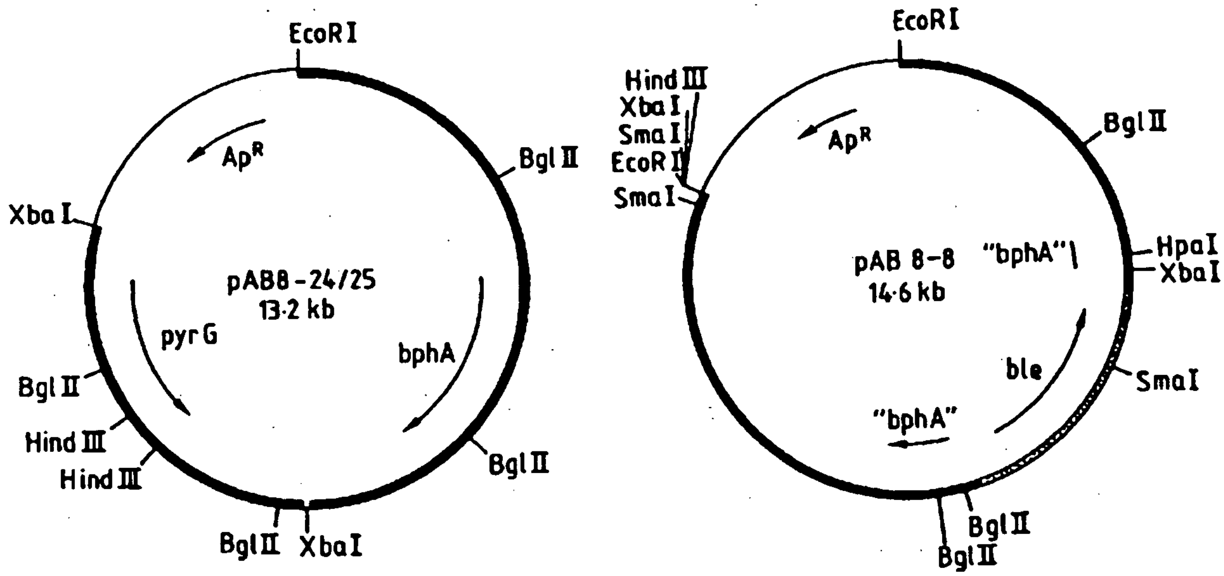


Fig. 1.2. Schematic representation of plasmid pAB8-8, which contains the disrupted *A. niger bphA* gene and the plasmids pAB8-24 and pAB8-25, which contain the *bphA* gene and, respectively, the wildtype or a mutant allele of the *pyrG* gene (Van Gorcom & Van den Hondel, 1988) of *A. niger* respectively. The disrupted *bphA* gene was obtained by replacing an *Eco*RV segment, located within the *bphA* gene, with the phleomycin resistance unit of pAN8-1. Thick line represents *A. niger* DNA, punctuated line *Streptoalloteichus hindustanus* DNA and thin line *E. coli* DNA; *ble*, phleomycin resistance gene; Ap^R, ampicillin resistance gene; '*bphA*', 5'- or 3'-terminal part of the *bphA* gene. Arrows indicate the direction of transcription.

analysis of the promoter region of the *Aspergillus* genes *gpdA*, *niaD* and *niiA* will be described. The third example deals with a study of the influence of different signal sequences on the efficiency of production of prochymosin in *A. niger*.

Cloning of a functional bphA *gene of* A. niger

Benzoate is metabolized by *A. niger* in a series of steps of which the first is *p*-hydroxylation of the aromatic ring of benzoate, carried out by benzoate-*p*-hydroxylase (BPH). Several mutants, disturbed in BPH activity, have been isolated (Boschloo & Bos, in preparation). These mutations were shown to belong to one complementation group, therefore the mutation was named *bphA*.

A cosmid clone, pAB8-1, containing the putative *bphA* gene, was isolated by differential hybridiaztion techniques. The gene was localized on a 6·2 kb *Eco*RI–*Pvu*II fragment, which was subcloned in pUC19, resulting in pAB8-22 (Van Gorcom *et al.*, 1990). Introduction of pAB8-1 or pAB8-22 DNA into an *A. niger bphA* mutant resulted in the restoration

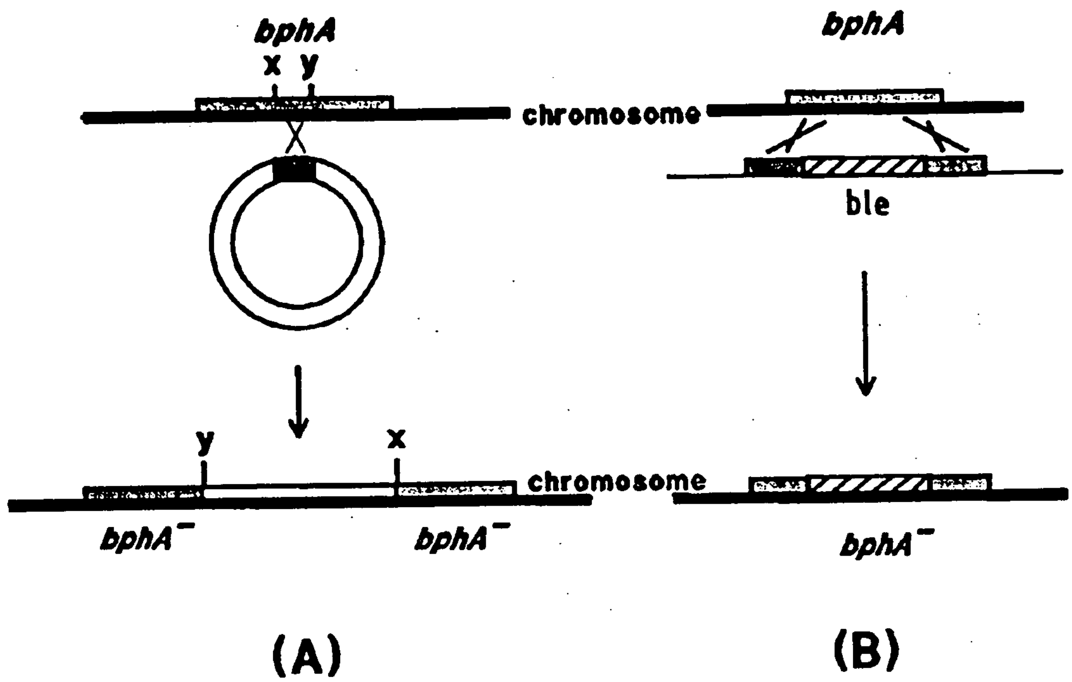


Fig. 1.3. Strategies to disrupt the *bphA* gene. Thin lines represent plasmid DNA and thick line chromosomal DNA. Shaded boxes represent the *bphA* gene or part of the gene. Hatched boxes represent the phleomycin resistance unit. Two restriction sites within the *bphA* gene are indicated with X and Y. (A) Disruption of the *bphA* gene by transformation with a plasmid which contains an internal restriction fragment. The recombination event shown results in formation of a duplication of *bphA* with the leftward copy lacking the 3' end of the gene and the rightward copy lacking the 5' end. (B) Disruption of the *bphA* gene by transformation with a linear fragment which contains a mutant allele of the *bphA* gene obtained by replacing an internal fragment with the phleomycin resistance unit. Recombination between the rightward and leftward homologous regions of the DNA fragment and the corresponding chromosomal regions results in a gene replacement of the wildtype gene with the mutant (disrupted) *bphA* allele.

---

of the ability to grow on benzoate, suggesting that the DNA fragment contained the *bphA* gene.

Although remote, it cannot be completely excluded that a suppressor of the *bphA* mutation had been cloned. One approach to exclude this possibility is to disrupt the cloned gene, replace the chromosomal gene by the disrupted equivalent and test for the inability to grow on benzoate.

Two methods regularly used in gene-disruption experiments are indicated in Fig. 1.3. In both cases the disruption vector contains a non-functional copy of the chromosomal gene to be disrupted. The method indicated in Fig. 1.3A requires knowledge about the exact position of the gene in the cloned fragment, whereas for the method indicated in Fig. 1.3B this is not necessary. To obtain an *A. niger* strain in which the

bphA gene was disrupted, the method indicated in Fig. 1.3B was chosen. For the disruption experiment, plasmid pAB8-8 was constructed (Fig. 1.2) which contains the non-functional *bphA* gene. In this plasmid part of the *bphA* sequences has been replaced by the phleomycin resistance unit of pAN8-1 (Mattern, Punt & Van den Hondel, 1988). Transformation of *A. niger* wild type with the isolated *Eco*RI fragment of pAB8-8 resulted in a number of phleomycin resistant colonies. Southern blot analysis revealed that in about 10% of the transformants a gene replacement had occurred. Further analysis showed that these transformants were not able to grow on benzoate as carbon source. This result confirms that the *bphA* gene and not a suppressor gene had been cloned.

Further evidence for cloning of the benzoate-*p*-hydroxylate-encoding gene was obtained from the DNA sequence of the *bphA* gene. Sequence comparison showed that the *bphA* gene encoded a cytochrome P450 mono-oxygenase, as might be expected.

Another important issue was the question whether the cloned gene was a functional copy of the *bphA* gene. To answer this question it was necessary to prove that the *bphA* mutation was complemented by the product of the cloned gene. Therefore an *A. niger bph*$^-$ strain was transformed with a plasmid containing the cloned gene and transformants were isolated in which the plasmid was integrated at an ectopic locus. Growth of these transformants on benzoate would indicate that a functional gene had been cloned. To achieve ectopic integration, the *A. niger pyrG* selection marker was cloned into pAB8-22 resulting in plasmid pAB8-24 (Fig. 1.2). Van Hartingsveldt *et al.* (1987) previously had found that a vector containing this selection marker is integrated at the *pyrG* locus in about 50% of *A. niger* transformants. However, Southern analysis of 48 transformants, obtained with pAB8.24, revealed that none of these transformants contained a vector integrated at the *pyrG* locus. Further analysis indicated that in most transformants the vector was integrated at the *bphA* locus.

To overcome the problem of preferential integration at the *bphA* locus, a mutant allele of the *A. niger pyrG* gene (Van Gorcom & Van den Hondel, 1988) was cloned in pAB8-22, resulting in pAB8-25 (Fig. 1.2). This mutant allele was constructed by introduction of a frameshift mutation which inactivates the marker gene. Transformation with the mutant allele as selection marker can result in Pyr$^+$ transformants only through type I or type III integration events. Analysis by Southern blotting of transformants obtained with pAB8-25 revealed that 14 out of 32 contained a single copy of this plasmid integrated at the *pyrG* locus. These transformants also showed a restored ability to grow on benzoate, indicating that, indeed, a functional *bphA* gene had been cloned. As demonstrated by Southern

analysis the other transformants resulted from a gene replacement at the *pyrG* locus. As expected, these transformants could not grow on benzoate.

Vectors for analysis of expression signals from **Aspergillus** *genes*

In both fundamental and applied molecular biological research on filamentous fungi the unravelling of the mechanism of gene expression is a very important topic. Interesting biological processes, such as development, differentiation and carbon and nitrogen metabolism are regulated at the level of gene expression. A wealth of classical genetic information is available for these processes, but, until recently, hardly any molecular genetic research was carried out. To provide an easy way to assay the expression and regulation of various genes, we developed reporter vectors for filamentous fungi (Van Gorcom *et al.*, 1986; Van Gorcom & Van den Hondel, 1988; Roberts *et al.*, 1989). In these vectors the analysis of fungal expression signals can be carried out by fusion of these signals to the *E. coli* reporter genes, *lacZ* or *uidA* encoding β-galactosidase and β-glucuronidase, respectively. The products of these genes can be assayed both qualitatively and quantitatively with easy and sensitive methods. For proper analysis of expression signals, it is essential that integration of one copy of the expression unit can be achieved at a specific location on the chromosome of the recipient. To fulfil this requirement, homologous selection markers were introduced in these vectors. An even higher (relative) frequency of homologous integration could be obtained by using mutant selection markers. These mutant selection markers were constructed by introduction of a frameshift mutation which inactivates the marker gene. Thus, only intragenic recombination (Type I or III integration) between the mutant selection marker on the vector and the mutant allele in the genome will result in prototrophic transformants. Although the transformation frequency obtained with this type of marker is much reduced (about 10-100 fold), Southern analysis of only a few transformants is sufficient to identify transformants with a single copy at the locus chosen (Table 1.5). Also, linearisation of the vector with a restriction enzyme which cuts in the marker gene, increases the relative frequency of Type I integration (Table 1.5).

The promoters of the *gpdA* genes of both *A. niger* and *A. nidulans* were studied in *A. niger* with the use of one of these vectors (Fig. 1.4). Single copy transformants, obtained with the two p_{gpdA}-*lacZ* fusion constructs, were assayed for β-galactosidase activity. In both cases efficient β-galactosidase expression was obtained (Table 1.6), whereas in untransformed strains or strains transformed with pAB94-12 (vector without promoter sequences inserted) no significant β-galactosidase activity was detected.

Table 1.5. Results of Southern analysis of *A. nidulans* transformants obtained with pAN5-d1 and derivatives

Vector[1]	Transformation frequency[2]	Percentage of LacZ$^+$ transformants[3]	Type of integration[4]		
			A	B	C
pAN5.d1	20-40	60%	0/19	4/19	15/19
pAN5.d1$_{(BglII\ digest)}$	40-100	90%	1/10	5/10	4/10
pAN5.d1$_{BglII}$	0·1-1	40%	5/10	3/10	2/10

[1] Vector pAN5-d1 contains a p_{gpdA}-*LacZ* fusion and the wildtype *argB* gene as selection marker for *Aspergillus* transformation (Punt *et al.*, 1990). The vector contains a unique *Bgl*II site in the coding region of the *argB* gene. Analysis of the transformants obtained with pAN5-d1, with a *Bgl*II digest of pAN5-d1 and with pAN5-d1$_{BglII}$, in which the unique *Bgl*II site was filled in with PolIK resulting in a frame shift mutation in the *argB* gene (Punt *et al.*, 1990), was carried out. Vectors were introduced into *A. nidulans* ArgB$^-$ (*methG2*, *biA1*, *argB2*).

[2] Transformation frequency is given as transformants per μg of vector DNA.

[3] The percentage of LacZ$^+$ transformants was determined by plating transformants on agar plates containing XGal (van Gorcom *et al.*, 1985). In all cases both LacZ$^+$ and LacZ$^-$ transformants were observed. The LacZ$^-$ transformants probably arose from gene replacement events.

[4] Southern analysis of a number of LacZ$^+$ transformants was carried out. The transformants were classified in three categories; A, single copy integration of the vector at the *argB* locus; B, multiple copy (tandem) integration of vector molecules at the *argB* locus; C, ectopic integration, in some cases in combination with single or multiple copy homologous integration.

A. nidulans and *A. niger* are both very efficient in *A. niger*. Further analysis of the organisation of the expression signals of the *A. nidulans* gene *gpdA* with similar vectors developed for *A. nidulans* is in progress in our laboratory (Punt *et al.*, 1990).

Recent research has shown that many fungal genes involved in developmental and metabolic pathways are organised as gene clusters (Gurr, Unkles & Kinghorn, 1988). Frequently, these clustered genes are coordinately expressed from divergently transcribing intergenic promoter regions. For the analysis of such intergenic regions a twin reporter vector was developed (Fig. 1.5). The usefulness of this vector can be inferred from the functional analysis of the intergenic region between the *A. nidulans* nitrate reductase (*niaD*) and nitrite reductase (*niiA*) genes. As shown in Table 1.7, both nitrate induction and nitrogen metabolite (ammonium) repression is observed for the reporter genes. Thus, the

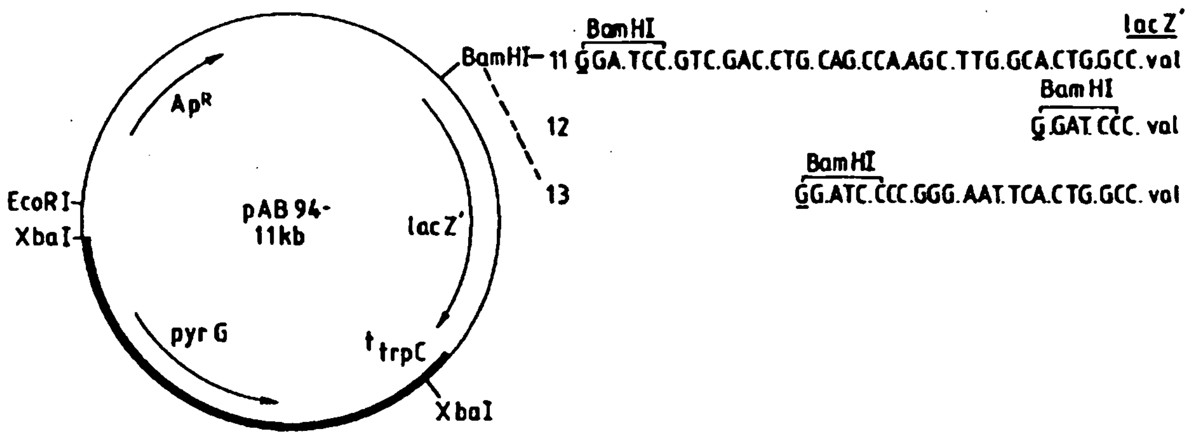


Fig. 1.4. Schematic representation of expression analysis vectors pAB94-11 to 13 for *A. niger* (Van Gorcom & Van den Hondel, 1988). The different vectors contain a unique *Bam*HI site in one of the three reading frames in front of the *lacZ'* gene (the protein coding region of the *E. coli lacZ* gene lacking the first eight codons). Thick line represents *A. niger* DNA (*Xba* I fragment) and *A. nidulans* DNA. Thin line represents *E. coli* DNA; t_{trpC}, terminator region of the *trpC* gene; Ap^R, ampicillin resistance gene; *pyrG*, mutant allele of the *A. niger pyrG* gene. Arrows indicate the direction of transcription.

Table 1.6. β-Galactosidase expression in *A. niger* transformants containing p_{gpdA}-*lacZ* fusion genes

Strain[1]	p_{gpdA}	βGAL activity[2]
AB4-1 [pAB94-53] 4	*A. niger*	8570
6		8380
7		7770
AB4-1 [pAB94-121] 4	*A. nidulans*	5160
13		5480
17		5350
AB4-1	–	<10

[1] Vectors pAB94-53 and pAB94-121, derivatives of pAB94-11/12/13, containing the promoter region of the *gpdA* gene of *A. niger* and *A. nidulans*, respectively, fused to the *LacZ* gene, were introduced into *A. niger* AB4-1 (*cspA*1, *pyrG*). Transformants with a single copy of the vector integrated at the *pyrG* locus were identified by Southern analysis.

[2] Enzyme activity is given in units (mg protein)$^{-1}$ and was measured as described

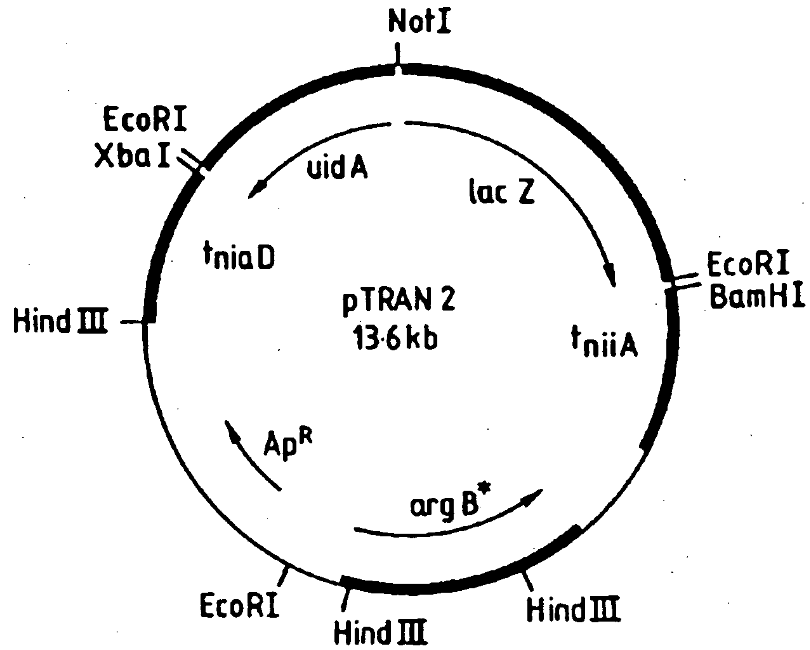


Fig. 1.5. Schematic representation of the twin reporter vector pTRAN2. Thin line represent pBR322 DNA; thick lines, *A. nidulans* DNA and *E. coli* DNA (*Eco*RI fragment) which contains the coding region of the *E. coli* genes *lacZ* and *uidA* both without translation initiation codon. A unique *Not*I site is placed between these genes; t_{niaD}, terminator region of the *niaD* gene; t_{niiA}, terminator region of the *niiA* gene; Ap^R, ampicillin resistance gene; *argB*$^+$, mutant allele of the *A. nidulans argB* gene containing a frameshift mutation. Arrows indicate the direction of transcription.

expression of the reporter genes *lacZ* and *uidA* faithfully represents the regulated gene expression of the genes *niaD* and *niiA* (Cove, 1979).

Expression of prochymosin

Several filamentous fungi are able to produce large amounts of extracellular proteins. Due to this property, several groups, including ours, are carrying out research to evaluate the potential of these strains for the production of heterologous proteins. One of the questions we addressed in our research on expression and secretion of heterologous, extracellular proteins in *A. niger* is the influence of different signal sequences on the efficiency of protein production/secretion. To answer this question experiments were performed to analyze the production of prochymosin with four different gene fusions (van Hartingsveldt *et al*., 1990). These fusions were placed under the control of the expression signals of the *A. niger* glucoamylase (*glaA*) gene. To facilitate proper comparison, transformants containing a single copy of the expression unit integrated at the *glaA* locus were isolated. For this purpose four different prochymosin expression vectors, pAB64-72 to pAB64-75 (Fig. 1.6), were used. Transformation of *A. niger* with *Hin*dIII-linearised pAB64-72 to 75 resulted in a number of hygromycin B resistant transformants. Southern analysis demonstrated

Table 1.7. Expression of the reporter genes in pTRAN2-1A transformants

	Relative enzyme activities[2]			
Strain[1]:	G324		SAA1012	
	βGUS	βGAL	βGUS	βGAL
proline	10	20	320	180
nitrate+proline	100	100	490	260
nitrate+ammonium	20	40	10	20
ammonium	2	10	2	4

[1] Vector pTRAN2-1A, a derivative of pTRAN2 (Fig. 1.4), containing the *A. nidulans niaD–niiA* intergenic promoter region was introduced in *A. nidulans* strains G324 (*wA3*, *yA2*, *methH2*, *argB2*, *galA1*, *sC12*, *ivoA1*) and SAA1012 (*fwA1*, *yA2*, *methH2*, *pabaA1*, *argB2*, *niiA-niaD* Δ509). Single copy transformants were identified by Southern analysis. In all cases the β-glucuronidase (βGUS) expression is a result of the activity of the expression signals of the *niaD* gene, and the β-galactosidase (βGAL) expression results from *niiA* gene expression signals.

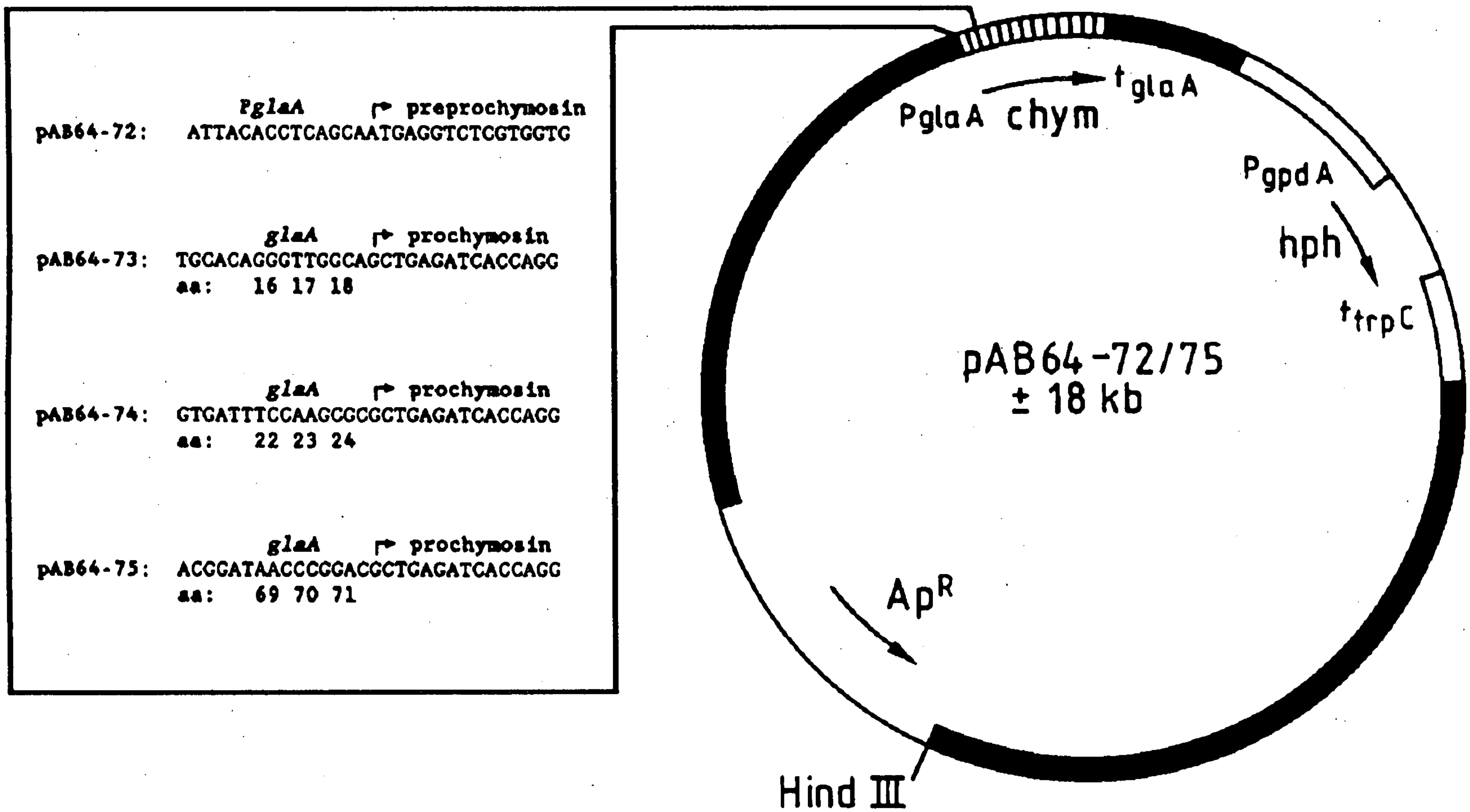
[2] Mycelial extracts were prepared from cells cultivated for 16 to 18 h in minimal growth medium with appropriate supplements and 10 mM of the indicated nitrogen sources. The enzyme activities were determined as described previously (Van Gorcom *et al.*, 1985; Roberts *et al.*, 1989) and are expressed relative to the activities of the G324[pTRAN2-1A] transformants induced with nitrate (= 100). In a representative experiment specific activities of 80 nmol *p*-nitrophenol min^{-1} (mg protein)$^{-1}$ for βGUS and 310 nmol *o*-nitrophenol min^{-1} (mg protein)$^{-1}$ for βGAL were found for the G324[pTRAN2-1A] transformants induced with nitrate.

Table 1.8. Analysis of prochymosin production in *A. niger*

Strain[1]	Signal peptide	Western[2] ($\mu g\ ml^{-1}$)	MCA[2] ($U\ ml^{-1}$)
AB64-72	signal sequence of prochymosin	6·2	8·6
AB64-73	signal sequence of *glaA*	11·3	19·5
AB64-74	signal sequence of *glaA* + 6 additional amino acids	4·1	3·2
AB74-75	signal sequence of *glaA* + 53 additional amino acids	10·2	19·1

[1] Vectors pAB64-72 to 75, linearised by cutting with *Hind*III, were introduced into *A. niger*. Transformants in which the *glaA* gene is replaced by the prochymosin fusion-genes, were identified by Southern hybridization (Van Hartingsveldt *et al.*, 1990).

[2] Medium samples from cells cultivated for 24 h in induction medium, were analyzed for the presence of prochymosin by Western blotting (Western) and



ɡ. 1.6. Schematic representation of four different prochymosin-expression vectors pAB64-72 to 75 (Van Hartingsveldt *et al.*, 90). Sequences at the glucoamylase-preprochymosin junction are indicated at the lefthand side. Thick lines represent *A. niger* NA from the *glaA* locus, open lines *A. nidulans* DNA, hatched lines preprochymosin DNA and thin lines *E. coli* DNA; p_{gpdA}, omoter region of the *gpdA* gene; p_{glaA}, promoter region of the *glaA* gene; t_{trpC}, terminator region of the *trpC* gene; t_{glaA}, terminator gion of the *glaA* gene; *hph*, hygromycin resistance gene; Ap^R, ampicillin resistance gene; *chym*, calf preprochymosin. Arrows dicate the direction of transcription.

that in about 10% of these transformants the resident *glaA* gene was replaced by the expression/secretion unit. Similar results were obtained with circular vector DNA, though with a three- to five-fold lower frequency.

Transformants which contained one copy of the expression/secretion unit were analyzed for prochymosin production 24 h after induction of the *glaA* promoter with starch (Van Hartingsveldt *et al.*, 1990). As shown in Table 1.8 similar levels of prochymosin were produced with pAB64-73 (18 amino acids of *glaA*) and pAB64-75 (71 amino acids of *glaA*). With 24 amino acids of *glaA* in front of prochymosin, or with the signal sequence of prochymosin itself, a lower production level was observed.

Although the reasons for the observed differences are obscure, our results clearly demonstrate that gene fusions containing different 5' sequences influence the production level of prochymosin.

Conclusions

During the last few years the development of gene transfer systems has been described for more than fifty fungal species. Transformation of most species could be achieved with heterologous auxotrophic markers or dominant selectable markers. Usually the marker gene is expressed from fungal, mainly *A. nidulans*, expression signals which were shown to be functional in most fungal species.

A number of strategies are now available for the development of gene transfer systems for hitherto poorly characterized fungal species. As illustrated in the first part of this chapter, these strategies comprise the following aspects. Firstly, methods for the introduction of vector DNA (Table 1.1). Secondly, a large number of auxotrophic and dominant selectable markers (Tables 1.2 to 1.4). Thirdly, efficient strategies for the isolation of auxotrophic mutant strains.

The main purpose for the development of gene transfer systems is application of these systems for molecular genetic studies. In the second part of this chapter several applications were illustrated with examples taken from research carried out in our laboratory. Genetic manipulation experiments were carried out (a) to disrupt the *bphA* gene of *A. niger*; (b) to analyze expression signals in *A. nidulans* and *A. niger*; (c) to direct expression-analysis vectors at specific sites of the genome such as the *argB* locus of *A. nidulans* or the *pyrG* locus of *A. niger* and (d) to perform gene replacement experiments in which the *glaA* gene of *A. niger* was replaced by chimeric prochymosin genes. These examples, as well as others described in the recent literature, indicate that most strategies and tools for genetic manipulation in filamentous fungi are now available, especially for *A. nidulans*, *A. niger* and *Neurospora crassa*. Extensive molecular genetic

studies of many interesting biological processes occurring in filamentous fungi can now be carried out using these strategies and tools.

Acknowledgements We wish to thank Peter Pouwels for critical reading of the manuscript and fruitful discussions. Gerry van Oossanen is acknowledged for typing the manuscript. Part of the work described in this chapter was financially supported by Gist-brocades, Delft, The Netherlands, DSM research, Geleen, The Netherlands and training grants from the EC.

References

Akins, R. A. & Lambowitz, A. M. (1985). General method for cloning *Neurospora crassa* nuclear genes by complementation of mutants. *Molecular and Cellular Biology* **5**, 2272-2278.

Armaleo, D., Ye, G. -N., Klein, T. M., Shark, K. B., Sanford, J. C. & Johnston, S. A. (1990). Biolistic nuclear transformation of *Saccharomyces cerevisiae* and other fungi. *Current Genetics* **17**, 97-103.

Avalos, J., Geever, R. F. & Case, M. E. (1989). Bialaphos resistance as a dominant selectable marker in *Neurospora crassa*. *Current Genetics* **16**, 369-372.

Ballance, D. J., Buxton, F. P. & Turner, G. (1983). Transformation of *Aspergillus nidulans* by the orotidine-5'-phosphate decarboxylase gene of *Neurospora crassa*. *Biochemical and Biophysical Research Communications* **112**, 284-289.

Ballance, D. J. & Turner, G. (1985). Development of a high frequency transforming vector for *Aspergillus nidulans*. *Gene* **36**, 321-331.

Ballance, D. J. & Turner, G. (1986). Gene cloning in *Aspergillus nidulans*: isolation of the isocitrate lyase gene (*acuD*). *Molecular and General Genetics* **202**, 271-275.

Banks, G. R. (1983). Transformation of *Ustilago maydis* by a plasmid containing yeast 2 micron DNA. *Current Genetics* **7**, 73-77.

Banks, G. R. & Taylor, S. Y. (1988). Cloning of the *PYR3* gene of *Ustilago maydis* and its use in DNA transformation. *Molecular and Cellular Biology* **8**, 5417-5424.

Bégueret, J., Razanamparany, V., Perrot, M. & Barreau, C. (1984). Cloning gene *ura5* for the orotidylic acid pyrophosphorylase of the filamentous fungus *Podospora anserina*: transformation of protoplasts. *Gene* **32**, 487-492.

Bej, A. K. & Perlin, M. H. (1989). A high efficiency transformation system for the basidiomycete *Ustilago maydis* employing hygromycin resistance and lithium-acetate treatment. *Gene* **80**, 171-176.

Binninger, O. M., Skrzynia, C., Pukkila, P. J. & Casselton, L. A. (1987). DNA-mediated transformation of the basidiomycete *Coprinus cinereus*. *EMBO Journal* **6**, 835-840.

Blakemore, E. J. A., Dobson, M. J., Hocart, M. J., Lucas, J. A. & Peberdy, J. F. (1989). Transformation of *Pseudocercosporella herpotrichoides* using heterologous genes. *Current Genetics* **16**, 177-180.

Brygoo, Y. & Debuchy, R. (1985). Transformation by integration in *Podospora anserina*.I. Methodology and phenomenology. *Molecular and General Genetics* **200**, 128-131.

Bull, J. H., Smith, D. J. & Turner, G. (1988). Transformation of *Penicillium chrysogenum* with a dominant selectable marker. *Current Genetics* **13**, 377-382.

Buxton, F. P. & Radford, A. (1984). The transformation of mycelial spheroplasts of *Neurospora crassa* and the attempted isolation of an autonomous replicator. *Molecular and General Genetics* **196**, 339-344.

Buxton, F. P., Gwynne, D. I. & Davies, R. W. (1985). Transformation of *Aspergillus niger* using the *argB* gene of *Aspergillus nidulans*. *Gene* **37**, 207-214.

Carramolino, L., Lozano, M., Pérez-Aranda, A., Rubio, V. & Sanchez, E. (1989). Transformation of *Penicillium chrysogenum* to sulfonamide resistance. *Gene* **77**, 31-38.

Case, M. E., Schweizer, M., Kushner, S. R. & Giles, N. H. (1979). Efficient transformation of *Neurospora crassa* by utilizing hybrid plasmid DNA. *Proceedings of the National Academy of Sciences, USA* **76**, 5259-5263.

Casselton, L. A. & De La Fuente Herce, A. (1989). Heterologous gene expression in the basidiomycete fungus *Coprinus cinereus*. *Current Genetics* **16**, 35-40.

Churchill, A. C. L., Guffetti, L. M., Hansen, D. R., Van Etten, H. D. & Van Alfen, N. K. (1990). Transformation of the fungal pathogen *Cryphonectria parasitica* with a variety of heterologous plasmids. *Current Genetics* **17**, 25-31.

Comino, A., Kolar, M., Schwab, H. & Socic, H. (1989). Heterologous transformation of *Claviceps purpurea*. *Biotechnology Letters* **11**, 389-392.

Cooley, R. N., Shaw, R. K, Franklin, F. C. H. & Caten, C. E. (1988). Transformation of the phytopathogenic fungus *Septoria nodorum* to hygromycin B resistance. *Current Genetics* **13**, 383-390.

Cooley, R. N. & Caten, C. E. (1989). Cloning and characterization of the β-tubulin gene and determination of benomyl resistance in *Septoria nodorum*. In *Proceedings, EMBO Alko Workshop on Molecular Biology of Filamentous Fungi*, vol. 6, (ed. H. Nevailainen & M. Penttilä), pp. 207-216. Foundation of Biotechnical and Industrial Fermentation Research: Helsinki.

Cove, D. J. (1979). Genetic studies of nitrate assimilation in *Aspergillus nidulans*. *Biological Reviews* **54**, 291-327.

Da Silva, A. J. F., Whittington, H., Clements, J., Roberts, C. & Hawkins, A. R. (1986). Sequence analysis and transformation by the catabolic 3-dehydroquinase (*QUTE*) gene from *Aspergillus nidulans*. *Biochemical Journal* **240**, 481-488.

De Graaff, L., Van den Broek, H. & Visser, J. (1988). Isolation and transformation of the pyruvate kinase gene of *Aspergillus nidulans*. *Current Genetics* **13**, 315-321.

De Ruiter-Jacobs, Y. M. J. T., Broekhuijsen, M., Campbell, E. I., Unkles, S. E., Kinghorn, J. R., Contreras, R., Pouwels, P. H. & van den Hondel, C. A. M. J. J. (1989). A gene transfer system based on the homologous *pyrG* gene and efficient expression of bacterial genes in *Aspergillus oryzae*. *Current Genetics* **16**, 159-163.

Durrens, P., Green, P. M., Arst, H. N. & Scazzocchio, C. (1986). Heterologous insertion of transforming DNA and generation of new deletions associated with transformation in *Aspergillus nidulans*. *Molecular and General Genetics* **203**, 544-549.

Farman, M. L. & Oliver, R. P. (1988). The transformation of protoplasts of *Leptosphaeria maculans* to hygromycin B resistance. *Current Genetics* **13**, 327-330.

Fincham, J. R. S. (1989). Transformation in fungi. *Microbiological Reviews* **53**, 148-170.

Glumoff, V., Käppeli, O., Fiechken, A. & Reiser, J. (1989). Genetic transformation of the filamentous yeast *Trichosporon cutaneum* using dominant selection markers. *Gene* **84**, 311-318.

Goldman, G. H., Van Montagu, M. & Herrera-Estrella, A. (1990). Transformation of *Trichoderma harzianum* by high-voltage electric pulse. *Current Genetics* **17**, 169-174.

Goosen, T., Bloemheuvel, G., Gysler, C., De Bie, D. A., Van den Broek, H. W. J. & Swart, K. (1987). Transformation of *Aspergillus niger* using the homologous orotidine-5'-phosphate-decarboxylase gene. *Current Genetics* **11**, 499-503.

Goosen, T., van Engelenburg, F., Debets, F., Swart, K., Bos, K. & Van den Broek, H. (1989). Tryptophan auxotrophic mutants in *Aspergillus niger*. Inactivation of the *trpC* gene by cotransformation mutagenesis. *Molecular and General Genetics* **219**, 282-288.

Goosen, T., Bos, C. J. & Van den Broek, H. W. J. (1991). Transformation and gene manipulation in filamentous fungi: an overview. In *Handbook of Applied Mycology* (Fungal Biotechnology volume 4), (ed. D. K. Arora, K. G. Mukerji & R. P. Elander), in press. M. Dekker: New York.

Goyon, C. & Faugeron, G. (1989). Targeted transformation of *Ascobolus immersus* and *de novo* methylation of the resulting duplicated DNA sequences. *Molecular and Cellular Biology* **9**, 2818-2827.

Gurr, S. J., Unkles, S. E. & Kinghorn, J. R. (1988). The structure and organisation of nuclear genes of filamentous fungi. In *Gene Structure in Eukaryotic Microbes*, Society of General Microbiology Special Publication volume 23, (ed. J. R. Kinghorn), pp. 93-139. IRL Press: Oxford.

Hargreaves, J. A. & Turner, G. (1989). Isolation of the acetyl-CoA synthase gene from the corn smut pathogen, *Ustilago maydis*. *Journal of General Microbiology* **135**, 2675-2678.

Henson, J. M., Blake, N. K. & Pilgeram, A. L. (1988). Transformation of *Gaeumannomyces graminis* to benomyl resistance. *Current Genetics* **14**, 113-117.

Herrera-Estrella, A., Goldman, G. H. & Van Montagu, M. (1990). High efficiency transformation system for the biocontrol agents *Trichoderma* spp. *Molecular Microbiology* **4**, 839-843.

Horng, J. S., Linz, J. E. & Pestka, J. J. (1989). Cloning and characterization of the *trpC* gene from an aflatoxigenic strain of *Aspergillus parasiticus*. *Applied and Environmental Microbiology* **55**, 2561-2568.

Iimura, Y., Gomi, K., Uzu, H. & Hara, S. (1987). Transformation of *Aspergillus oryzae* through plasmid-mediated complementation of the methionine-auxotrophic mutation. *Agricultural and Biological Chemistry* **51**, 323-328.

John, M. A. & Peberdy, J. F. (1984). Transformation of *Aspergillus nidulans* using the *argB* gene. *Enzyme Microbial Technology* **6**, 386-389.

Kelly, J. M. & Hynes, M. J. (1985). Transformation of *Aspergillus niger* by the *amdS* gene of *Aspergillus nidulans*. *EMBO Journal* **4**, 475-479.

Kim, S. Y. & Marzluf, G. A. (1988). Transformation of *Neurospora crassa* with the *trp-1* gene and the effect of host strain upon the fate of the transforming DNA. *Current Genetics* **13**, 65-70.

Kinsey, J. A. & Rambosek, J. A. (1984). Transformation of *Neurospora crassa* with the cloned *am* (glutamate dehydrogenase) gene. *Molecular and Cellular Biology* **4**, 117-122.

Kolar, M., Punt, P. J., Van den Hondel, C. A. M. J. J. & Schwab, H. (1988). Transformation of *Penicillium chrysogenum* using dominant selection markers and expression of an *Escherichia coli lacZ* fusion gene. *Gene* 62, 127-134.

Kornegay, M. A. J. R., Pribnow, D. & Gold, M. H. (1989). Transformation by complementation of an adenine auxotroph of the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. *Applied and Environmental Microbiology* **55**, 406-411.

Kos, A., Kuijvenhoven, J., Wernars, K., Bos, C. J., Van den Broek, H. W. J., Pouwels, P. H. & Van den Hondel, C. A. M. J. J. (1985). Isolation and characterization of the *Aspergillus niger trpC* gene. *Gene* **39**, 231-238.

Kronstad, J. W., Wang, J., Covert, S. F., Holden, D. W., McKnight, G. L. & Leong, S. A. (1989). Isolation of metabolic genes and demonstration of gene disruption in the phytopathogenic fungus *Ustilago maydis*. *Gene* **79**, 97-106.

Malardier, L., Daboussi, M. J., Julien, J., Roussel, F., Scazzocchio, C. & Brygoo, Y. (1989). Cloning of the nitrate reductase gene (*niaD*) of *Aspergillus nidulans* and its use for transformation of *Fusarium oxysporum*. *Gene* **78**, 147-156.

Mattern, I. E., Punt, P. J. & Van den Hondel, C. A. M. J. J. (1988). A vector of *Aspergillus* transformation conferring phleomycin resistance. *Fungal Genetics Newsletter* **35**, 25.

May, G. S., Gambino, J., Weatherbee, J. A. & Morris, N. R. (1985). Identification and functional analysis of β-tubulin genes by site-specific integrative transformation in *Aspergillus nidulans*. *Journal of Cell Biology* **100**, 712-718.

May, G. S., Waring, R. B., Osmani, S. A., Morris, N. R. & Denison, S. H. (1989). The coming of age of molecular biology in *Aspergillus nidulans*. In *Proceedings, EMBO Alko Workshop on Molecular Biology of Filamentous Fungi*, vol. 6, (ed. H. Nevailainen & M. Penttilä), pp. 11-20. Foundation of Biotechnical and Industrial Fermentation Research: Helsinki.

Mooibroek, H., Kuipers, A. G. J., Sietsma, J. H., Punt, P. J. & Wessels, J. G. H. (1990). Introduction of hygromycin B resistance into *Schizophyllum commune*: preferential methylation of donor DNA. *Molecular and General Genetics* **222**, 41-48.

Mullaney, E. J., Punt, P. J. & Van den Hondel, C. A. M. J. J. (1988). DNA mediated transformation of *Aspergillus ficuum*. *Applied and Microbial Biotechnology* **28**, 451-454.

Munoz-Rivas, A., Specht, C. A., Drummond, B. J., Froeliger, E., Novotny, C. P. & Ullrich, R. C. (1986). Transformation of the basidiomycete, *Schizophyllum commune*. *Molecular and General Genetics* **205**, 103-106.

Oakley, B. R., Rinehart, J. E., Mitchell, B. L., Oakley, C. E., Carmona, C., Gray, G. L & May, G. S. (1987a). Cloning, mapping and molecular analysis of the *pyrG* orotidine-5'-phosphate decarboxylase gene of *Aspergillus nidulans*. *Gene* **61**, 385-399.

Oakley, C. E., Weil, C. F., Kretz, P. L. & Oakley, B. R. (1987b). Cloning of the *riboB* locus of *Aspergillus nidulans*. *Gene* **53**, 293-298.

Oliver, R. P., Roberts, I. N., Harling, R., Kenyon, L., Punt, P. J., Dingemanse, M. A. & Van den Hondel, C. A. M. J. J. (1987). Transformation of *Fulvia fulva*, a fungal pathogen of tomato, to hygromycin B resistance. *Current Genetics* **12**, 231-233.

Orbach, M. J., Porro, E. B. & Yanofsky, C. (1986). Cloning and characterization of the gene for β-tubulin from a benomyl-resistant mutant of *Neurospora crassa* and its use as a dominant selectable marker. *Molecular and Cellular Biology* **6**, 2452-2461.

Osiewacz, H. D. & Weber, A. (1989). DNA mediated transformation of the filamentous fungus *Curvularia lunata* using a dominant selectable marker. *Applied and Microbial Biotechnology* **30**, 375-380.

Paietta, J. & Marzluf, G. A. (1985). Plasmid recovery from transformants and the isolation of chromosomal DNA segments improving plasmid replication in *Neurospora crassa*. *Current Genetics* **9**, 383-388.

Panaccione, D. G., McKiernan, M. & Hanau, R. M. (1988). *Colletotrichum graminicola* transformed with homologous and heterologous benomyl-resistance genes retains expected pathogenicity to corn. *Molecular plant-Microbe Interactions* **1**, 113-120.

Peberdy, J. F. (1989). Fungi without coats–protoplasts as tools for mycological research. *Mycological Research* **93**, 1-20.

Perrot, M., Barreau, C. & Bégueret, J. (1987). Nonintegrative transformation in the filamentous fungus *Podospora anserina*: stabilization of a linear vector by the chromosomal ends of *Tetrahymena thermophila*. *Molecular and Cellular Biology* **7**, 1725-1730.

Picknett, T. M., Saunders, G., Ford, P. & Holt, G. (1987). Development of a gene transfer system for *Penicillium chrysogenum*. *Current Genetics* **12**, 449-455.

Punt, P. J., Oliver, R. P., Dingemanse, M. A., Pouwels, P. H. & Van den Hondel, C. A. M. J. J. (1987). Transformation of *Aspergillus* based on the hygromycin B resistance marker from *Escherichia coli*. *Gene* **56**, 117-124.

Punt, P. J., Dingemanse, M. A., Kuyvenhoven, A., Soede, R. D. M., Pouwels, P. H. & Van den Hondel, C. A. M. J. J. (1990). Functional elements in the promoter region of the *Aspergillus nidulans gpdA* gene, encoding glyceraldehyde-3-phosphate dehydrogenase. *Gene* **93**, 101-109.

Queener, S. W., Ingolia, T. D., Skatrud, P. L., Chapman, J. L. & Kaster, K. R. (1985). A system for genetic transformation of *Cephalosporium acremonium*. In *Microbiology – 1985,* (ed. L. Leive), pp. 468-472.American Society of Microbiology: Washington, DC.

Radford, A., Pope, S., Scazi, A., Fraser, M. J. & Parish, J. H. (1981). Liposome-mediated genetic transformation of *Neurospora crassa*. *Molecular and General Genetics* **184**, 567-569.

Revuelta, J. L. & Jayaram, M. (1986). Transformation of *Phycomyces blakesleeanus* to G-418 resistance by an autonomously replicating plasmid. *Proceedings of the National Academy of Sciences, USA* **83**, 7344-7347.

Roberts, I. N., Oliver R. P., Punt P. J. & Van den Hondel, C. A. M. J. J. (1989). Expression of the *Escherichia coli* β-glucuronidase gene in industrial and phytopathogenic filamentous fungi. *Current Genetics* **15**, 177-180.

Samac, D. A. & Leong, S. A. (1989). Characterization of the termini of linear plasmids from *Nectria haematococcus* and their use in construction of an autonomously replicating transformation vector. *Current Genetics* **16**, 187-194.

Sánchez, E., Lozano, M., Rubio, V. & Penalva, M. A. (1987). Transformation in *Penicillium chrysogenum*. *Gene* **51**, 97-102.

Staben, C., Jensen, B., Singen, M., Pollock, J., Schechtman, M. & Kinsey, J. (1989). Use of a bacterial hygromycin B resistance gene as a dominant selectable marker in *Neurospora crassa* transformation. *Fungal Genetics Newsletter* **36**, 79-81.

Stinchcomb, D. T., Struhl, K. & Davis, R. W. (1979). Isolation and characterization of yeast chromosomal replicator. *Nature* **282**, 39-43.

Thomas, M. D. & Kenerly, C. M. (1989). Transformation of the mycoparasite *Gliocladium*. *Current Genetics* **15**, 415-420.

Tilburn, J., Scazzocchio, C., Taylor, G. G., Zabicky-Zissman, J. H., Lockington, R. A. & Davies, R. W. (1983). Transformation by integration in *Aspergillus nidulans*. *Gene* **26**, 205-221.

Timberlake, W. E. & Marshall, M. A. (1989). Genetic engineering of filamentous fungi. *Science* **244**, 1313- 1317.

Tsukuda, T., Carleton, S., Fotheringham, S. & Holloman, W. K. (1988). Isolation and characterization of an autonomously replicating sequence from *Ustilago maydis*. *Molecular and Cellular Biology* **8**, 3703-3709.

Turgeon, B. G., MacRae, W. D., Garber, R. C., Fink, G. R. & Yoder, O. C. (1986). A cloned tryptophan-synthesis gene from the Ascomycete *Cochliobolus heterostrophus* functions in *Escherichia coli*, yeast and *Aspergillus nidulans*. *Gene* **42**, 79-88.

Unkles, S. E., Campbell, E. I., Carrez, D., Grieve, C., Contreras, R., Fiers, W., Van den Hondel, C. A. M. J. J. & Kinghorn, J. R. (1989). Transformation of *Aspergillus niger* with the homologous nitrate reductase gene. *Gene* **78**, 157-166.

Unkles, S. E., Campbell, E. I., De Ruiter-Jacobs, Y. M. J. T., Broekhuijsen, M., Macro, J. A., Carrez, D., Contreras, R., Van den Hondel, C. A. M. J. J. & Kinghorn, J. R. (1989). The development of a homologous transformation system for *Aspergillus oryzae* based on the nitrate assimilation pathway: A convenient and general selection system for filamentous fungal transformation. *Molecular and General Genetics* **218**, 99-104.

Van Gorcom, R. F. M., Boschloo, J. G., Kuyvenhoven, A., Lange, J., Van Vark, A. J., Bos, C. J., Van Balken, J. A. M. & Van den Hondel, C. A. M. J. J. (1990). Isolation and molecular characterization of the benzoate-para-hydroxylase gene (*bphA*) of *A. niger* a member of a new gene-family of the cytochrome P450 superfamily. *Molecular and General Genetics* **223**, 192-197.

Van Gorcom, R. F. M., Pouwels, P. H., Goosen, T., Visser, J., Van den Broek, H. W. J., Hamer, J. E., Timberlake, W. E. & Van den Hondel, C. A. M. J. J. (1985). Expression of an *Escherichia coli* β-galactosidase fusion gene in *Aspergillus nidulans*. *Gene* **40**, 99-106.

Van Gorcom, R. F. M., Punt, P. J., Pouwels, P. H. & Van den Hondel, C. A. M. J. J. (1986). A system for the analysis of expression signals in *Aspergillus*. *Gene* **48**, 211-217.

Van Gorcom, R. F. M. & Van den Hondel, C. A. M. J. J. (1988). Expression analysis vectors for *Aspergillus niger*. *Nucleic Acid Research* **16**, 9052.

Van Hartingsveldt, W., Mattern, I. E., Van Zeijl, C. M. J., Pouwels, P. H. & Van den Hondel, C. A. M. J. J. (1987). Development of a homologous transformation system for *Aspergillus niger* based on the *pyrG* gene. *Molecular and General*

Van Hartingsveldt, W., Van den Hondel, C. A. M. J. J., Veenstra, A. E. & Van den Berg, J. A. (1990). Gene replacement as a tool for the construction of *Aspergillus* strains. European Patent Application 89202106.4.

Van Heeswijck, R. (1986). Autonomous replication of plasmids in *Mucor* transformants. *Carlsberg Research Communications* **51**, 433-443.

Van Heeswijck, R. & Roncere, M. I. G. (1984). High frequency transformation of *Mucor* with recombinant plasmid DNA. *Carlsberg Research Communications* **49**, 691-702.

Vollmer, S. J. & Yanofsky, C. (1986). Efficient cloning of genes of *Neurospora crassa*. *Proceedings of the National Academy of Sciences, USA* **83**, 4869-4873.

Ward, M., Kodama, K. H. & Wilson, L. J. (1989). Transformation of *Aspergillus awamori* and *Aspergillus niger* by electroporation. *Experimental Mycology* **13**, 289-293.

Ward, M., Wilkinson, B. & Turner, G. (1986). Transformation of *Aspergillus nidulans* with a cloned oligomycin resistant ATP synthase subunit 9 gene. *Molecular and General Genetics* **202**, 265-270.

Ward, M., Wilson, L. J., Carmona, C. L. & Turner, G. (1988). The *oliC3* gene of *Aspergillus niger*: isolation, sequence and use as a selectable marker for transformation. *Current Genetics* **14**, 37-42.

Wernars, N., Goosen, T., Wennekes, L. M. J., Visser, J., Bos, C. J., Van den Broek, H. W. J., Van Gorcom, R. F. M., Van den Hondel, C. A. M. J. J. & Pouwels, P. H. (1985). Gene amplification in *Aspergillus nidulans* by transformation with vectors containing the *amdS* gene. *Current Genetics* **9**, 361-368.

Whitehead, M. P., Unkles, S. E., Ramsden, M., Campbell, E. I., Gurr, S. J., Spence, D., Van den Hondel, C. A. M. J. J., Contreras, R. & Kinghorn, J. R. (1989). Transformation of a nitrate reductase deficient mutant of *Penicillium chrysogenum* with the corresponding *Aspergillus niger* and *A. nidulans niaD* genes. *Molecular and General Genetics* **216**, 408-411.

Wnendt, F., Jacobs, M. & Stahl, U. (1990). Transformation of *Aspergillus giganteus* to hygromycin B resistance. *Current Genetics* **17**, 21-24.

Wöstemeyer, J., Burmester, A. & Weigel, C. (1987). Neomycin resistance as a dominantly selectable marker for transformation of the zygomycete *Absidia glauca*. *Current Genetics* **12**, 625-627.

Yelton, M. M., Hamer, J. E. & Timberlake, W. E. (1984). Transformation of *Aspergillus nidulans* by using a *trpC* plasmid. *Proceedings of the National Academy of Sciences, USA* **81**, 1470-1474.

[redacted]
68

Gene Manipulations in Fungi

Edited by

J. W. Bennett

Department of Biology
Tulane University
New Orleans, Louisiana

Linda L. Lasure

Bioproduct Research
Miles Laboratories, Inc.
Elkhart, Indiana

1985

ACADEMIC PRESS, INC.
HARCOURT BRACE JOVANOVICH, PUBLISHERS
Orlando San Diego New York
Austin London Montreal Sydney
Tokyo Toronto

ACADEMIC PRESS, INC.
Orlando, Florida 32887

United Kingdom Edition published by
ACADEMIC PRESS INC. (LONDON) LTD.
24–28 Oval Road, London NW1 7DX

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 85-5979

ISBN 0–12–088640–5 (alk. paper)
ISBN 0–12–088641–3 (paperback: alk. paper)

PRINTED IN THE UNITED STATES OF AMERICA

85 86 87 88 9 8 7 6 5 4 3 2 1

Contents

Contributors xiii

Preface xvii

Acknowledgments xix

I Historical Perspective: Mutants to Models

1 From Auxotropic Mutants to DNA Sequences

John R. S. Fincham

I. The Establishment of the Paradigm: One Gene–One Polypeptide 3
II. Cluster Genes and Gene Clusters 10
III. Systems of Regulation of Gene Activity 13
IV. Cloning and Sequencing 22
References 28

2 Molecular Taxonomy of the Fungi

C. P. Kurtzman

I. Introduction 35
II. Nucleic Acid Isolation and Purification 37

III. DNA Base Composition 38
IV. DNA Relatedness 41
V. Mitochondrial DNA Relatedness 49
VI. Ribosomal RNA Relatedness 51
VII. Comparison of Relatedness from Nucleic Acid Studies with That Determined by Other Methodologies 53
References 56

3 Fungal Mitochondrial Genomes

Lawrence T. Grossman and Michael E. S. Hudspeth

I. Introduction 66
II. Physical Organization 66
III. Genes and the Genetic Code 70
IV. Gene Order 74
V. Transcription 76
VI. Pseudogenes 81
VII. Mitochondrial Plasmids 83
VIII. Effects of Alteration of mtDNA 84
IX. Nuclear–Cytoplasmic Interactions 88
References 91

4 Modeling the Environment for Gene Expression

Barbara E. Wright

I. Introduction 105
II. Modeling Metabolism under Steady-State Conditions 106
III. Modeling Metabolism during Product Accumulation 115
References 121

II Yeasts

5 *Saccharomyces cerevisiae* as a Paradigm for Modern Molecular Genetics of Fungi

Jasper Rine and Marian Carlson

I. Introduction 126
II. Neoclassical Genetics 127
III. Transformation 133
IV. Cloning a Gene 141
V. Manipulation of a Cloned Gene 141

VI. Other Uses of Cloned Genes 152
VII. An Agenda for Progress 153
References 155

6 Yeast Transformation

George Boguslawski

I. Introduction 161
II. Specific Transformation Systems in Yeast 167
III. Applications 178
IV. Conclusions 187
References 188

7 Use of the *LYS2* Gene for Gene Disruption, Gene Replacement, and Promoter Analysis in *Saccharomyces cerevisiae*

Debra A. Barnes and Jeremy Thorner

I. Introduction 197
II. Characterization of the *LYS2* Gene and Its Product 203
III. Genetic Manipulations Utilizing the *LYS2* Gene 210
IV. Conclusions and Prospectus 222
References 223

III Molds

8 Molecular Biology of the *qa* Gene Cluster of *Neurospora*

Layne Huiet and Mary Case

I. Introduction 229
II. Molecular Analysis of the *qa* Cluster 230
III. Transformation of the *qa* Gene Cluster in *Neurospora* 236
IV. Summary 242
References 242

9 *Neurospora* Plasmids

John A. Kinsey

I. Introduction 245
II. Naturally Occurring Plasmids 245

III. Replicating Plasmids Constructed in the Laboratory 248
IV. Overview and Prospects for the Future 253
References 256

10 Cloning and Transformation in *Aspergillus*

G. Turner and D. J. Ballance

I. Introduction 259
II. Transformation Methodology 260
III. Fate of Transforming DNA after Entry 265
IV. The Quest for Replicating Vectors 270
V. Development of Vectors 273
VI. Future Prospects 275
References 275

11 Expression of *Aspergillus* Genes in *Neurospora*

Richard L. Weiss, Diane Puetz, and Jan Cybis

I. Transformation of *Neurospora crassa* 280
II. Isolation of the *Aspergillus nidulans* Gene for Ornithine Carbamoyltransferase 281
III. Transformation of the *Neurospora crassa arg-12* Mutant .. 282
IV. Control of Gene Expression 288
V. Subcellular Localization of the Active Enzyme 289
VI. Conclusions 290
References 291

12 Gene Dosage Effects and Antibiotic Synthesis in Fungi

C. H. O'Donnell, A. Upshall, and K. D. Macdonald

I. Introduction 293
II. Methods for the Amplification of Genetic Material in Microorganisms 295
III. Studies with Disomic Strains of *Aspergillus nidulans* 297
References 305

13 Formal Genetics and Molecular Biology of the Control of Gene Expression in *Aspergillus nidulans*

Herbert N. Arst, Jr., and Claudio Scazzocchio

I. Formal Genetic Methodology of *Aspergillus nidulans* as Applied to the Study of Control Systems 310

II. The Metabolic Versatility of *A. nidulans* and Its Exploitation. . . . 311
III. Regulatory Genes 313
IV. Putative Receptor Sites 327
V. The Spatial Organization of Functionally Related Genes . . . 330
VI. At What Level Does Regulation of Gene Expression Occur? . 335
References 337

14 A Cloning Strategy in Filamentous Fungi

P. M. Green and Claudio Scazzocchio

Text 345
References 352

IV Applications

15 Primary Metabolism and Industrial Fermentations

Ramunas Bigelis

I. Primary Metabolites 358
II. Genetic Approaches to the Production of Primary Metabolites 359
III. Organic Acids 360
IV. Amino Acids 368
V. Polysaccharides 372
VI. Lipids 374
VII. Nucleotides and Nucleic Acid–Related Compounds 376
VIII. Vitamins 378
IX. Polyols 380
X. Ethanol 382
XI. The Promise of Biotechnology 383
References 385

16 Mitochondrial DNA for Gene Cloning in Eukaryotes

Paul Tudzynski and Karl Esser

I. Introduction 403
II. Fungal Plasmids 404
III. Cloning Vectors of Mitochondrial Origin 409
IV. Biotechnological Implications 412
References 413

17 Molecular Bases of Fungal Pathogenicity to Plants

O. C. Yoder and B. Gillian Turgeon

I. Introduction 417
II. Infection Structures 418
III. Cutinase 421
IV. Pisatin Demethylase 422
V. Toxins 424
VI. Cloning and Analysis of Pathologically Important Genes from Fungi 427
References 441

18 Morphogenesis and Dimorphism of *Mucor*

Ronald L. Cihlar

I. Introduction 449
II. Dimorphism of *Mucor* 450
III. Molecular Analysis of *Mucor* 457
IV. Perspectives—Molecular Genetics 462
References 464

19 Toward Gene Manipulations with Selected Human Fungal Pathogens

W. LaJean Chaffin

I. Introduction 469
II. *Cryptococcus neoformans* 471
III. *Histoplasma capsulatum* 473
IV. *Wangiella dermatitidis* 474
V. *Candida albicans* 476
VI. Concluding Remarks 486
References 487

20 Fungal Carbohydrases: Amylases and Cellulases

Bland S. Montenecourt and Douglas E. Eveleigh

I. Introduction 491
II. Fungal Amylases 494
III. Fungal Cellulases 501
IV. Summary and Outlook 505
References 508

V Postscript

21 Prospects for a Molecular Mycology

J. W. Bennett

I. Introduction 515
II. A Primer in Mycology 516
III. Molecular Mycology 519
References 526

Appendixes

I Fungal Taxonomy

Linda L. Lasure and J. W. Bennett

I. Introduction 531
II. Outline of Fungal Taxonomy 534
References 534

II Conventions for Gene Symbols

J. W. Bennett and Linda L. Lasure

I. *Aspergillus nidulans* 538
II. *Neurospora crassa* 539
III. *Saccharomyces cerevisiae* 540
IV. Other Fungi 542
References 542

Index 545

I

Fungal Taxonomy

LINDA L. LASURE
Bioproduct Research
Miles Laboratories, Inc.
Elkhart, Indiana

J. W. BENNETT
Department of Biology
Tulane University
New Orleans, Louisiana

I. Introduction 531
II. Outline of Fungal Taxonomy 534
References 534

I. INTRODUCTION

The purpose of this section is to provide a brief explanation of fungal taxonomy and a reference table for identifying major taxa. Remember that taxonomic schemes are neither static nor universally accepted. The one presented below follows Ainsworth (1971) and Ainsworth *et al.* (1973a,b). Other authorities may present quite different hierarchies and headings. Nomenclatural convention for fungi demands that subdivisions end in "-mycotina," classes in "-mycetes," orders in "-ales," and families in "-aceae." Depending on the authority and the scheme adopted, you may find the same group accorded differential rank. For example, the ascus-producing fungi may be viewed as a class, Ascomycetes, or as a subdivision, Ascomycotina.

If you are interested in exposure to other taxonomic arrangements and in learning more about mycology in general, consult one of the comprehensive, recent mycology texts such as Burnett (1968), Alexopoulos and Mims (1979), Ross (1979), or Moore-Landecker (1982).

The kingdom Fungi is divided here into two divisions. The Myxomycota, commonly called the "slime molds," are a varied group of organisms having a plasmodium at some point in their life cycle. One contemporary mycologist pointed out that "the very words *slime mold* reflect the confusion that has surrounded this group of organisms, because they are certainly not molds and they are not particularly slimy" (Ross, 1979, p. 178). A number of taxonomic questions remain unanswered as to whether the members of the Myxomycota really belong with the fungi.

Members of the division Eumycota, commonly called the "true fungi," usually have a filamentous or yeastlike form, and no plasmodium. Our scheme divides the group into five subdivisions. The Mastigomycotina and Zygomycotina constitute the "lower fungi"; the Ascomycotina, Basidiomycotina, and Deuteromycotina constitute the "higher fungi."

The lower fungi are distinguished by hyphae without cross-walls (nonseptate), the formation of asexual spores by cleavage of cytoplasm with sporangia, and include several groups that possess flagellated zoospores. For many years, the lower fungi were grouped together in a single class, the Phycomycetes. Phycomete means "algal fungus" and the name stems from the theory that these fungi were degenerate algae that had lost their chlorophyll. The term phycomycete no longer has official taxonomic status, but is still encountered in older texts and in works by authors who have not kept up with trends in fungal systematics. The classification we present here puts the lower fungi into two subdivisions, both of which encompass a diverse and composite group of organisms.

The subdivision Mastigomycotina includes species often identified with animals because of the defining characteristic of the group, motile spores. Many of these organisms are called water molds because of the prevalent aquatic growth habit.

Zygomycotina contains nonseptate fungi which lack a motile stage and are only rarely aquatic. Members of this subdivision exhibit gametangial fusion and zygospore formation.

Taxonomically, the higher fungi are easier to delineate. With the exception of the yeasts, they have septate hyphae and often produce elaborate fruiting bodies. They are divided here into three subdivisions: the Ascomycotina, the Basidiomycotina, and the Deuteromycotina. The Ascomycotina and Basidiomycotina are distinguished by their sexual spores; the Deuteromycotina reproduce entirely by asexual means.

The Ascomycotina form ascospores inside a specialized reproductive structure called an ascus. Two haploid nuclei fuse within the immature ascus and then the diploid fusion nucleus immediately undergoes meiosis, resulting in four haploid spores. One mitotic division usually ensues so that most members of the As-

comycotina have eight-spored asci. The retention of the products of meiosis within a single morphological structure has facilitated many elegant studies on chromosomal mechanisms of crossing-over. The three premier species for fungal genetics, *Aspergillus nidulans, Neurospora crassa,* and *Saccharomyces cerevisiae,* are all members of this group. Special features of fungal genetic analysis are discussed in detail by Esser and Kuenen, (1967), Burnett (1975), and Fincham *et al.* (1979).

The Basidiomycotina form sexual basidiospores on a basidium. Basidiospore formation closely resembles ascospore development, except that the spores are borne externally. Fusion of haploid nuclei results in a transient diploid that immediately undergoes meiosis to form four haploid basidiospores. An unusual cytological feature of the basidiomycete life cycle is the formation of a special binucleate cell called the dikaryon. This subdivision contains the majority of conspicuous, macroscopic fungi such as mushrooms, puffballs, and shelf fungi. It also contains the important plant pathogens known collectively as rusts and smuts.

The Deuteromycotina, or Fungi Imperfecti, are distinguished by the absence of any known sexual form. They reproduce largely by asexual conidiospores. Taxonomists consider this an ''artificial'' group, and often highlight this artificiality by using the prefix ''form'' with reference to the taxa within this subdivision (e.g., form-class, form-family, form-genus, form-species). Many species originally classified as imperfects are eventually shown to possess a sexual stage, usually within the Ascomycotina or, more rarely, within the Basidiomycotina. The sexual phase, also called the perfect stage or teleomorph, is given a separate name. The rules of botanical nomenclature specify that sexual names should have precedence over the asexual (also called imperfect or anamorphic) names. This creates both practical and philosophical problems. The genus names *Aspergillus, Penicillium,* and *Fusarium* are all imperfect epithets. According to the internationally adopted rules of nomenclature, any time a sexual stage is found for a member of one of these genera, the name of that species should be changed to that of the sexual form. For example, according to these rules, *Aspergillus nidulans* should be called *Emericella nidulans*. In practice, despite the fact that this species regularly forms ascospores, virtually everyone still calls it *Aspergillus nidulans*.

Many economically important fungi are classified in the Deuteromycotina. For more details about the taxonomy of *Aspergillus* see Raper and Fennell (1965); for *Fusarium* see Nelson *et al.* (1981); and for *Pencillium* see Pitt (1979) and Ramirez (1982). The majority of important human pathogens also belong to this group; see Rippon (1982). Finally, for a discussion of the issues and problems surrounding nomenclatural conventions in the Fungi Imperfecti, see Bennett (1985).

II. OUTLINE OF FUNGAL TAXONOMY

Kingdom: Fungi
Division: Myxomycota (plasmodium or pseudoplasmodium present)
Class: Acrasiomycetes ("cellular slime molds")
Example: *Dictyostelium*
Class: Myxomycetes ("acellular slime molds")
Example: *Physarum*
Division: Eumycota (assimilative phase typically filamentous or yeastlike)
Subdivision: Mastigomycotina (nonseptate mycelium, motile spores)
Examples: *Achlya, Allomyces, Blastocladiella, Phythium, Phytophthora, Saprolegnia*
Subdivision: Zygomycotina (nonseptate mycelium, zygospores)
Examples: *Absidia, Blakeslea, Mortierella, Mucor, Pilobolus, Rhizopus*
Subdivision: Ascomycotina ("sac fungi"; septate mycelium or yeast: sexual spores borne in an ascus)
Examples: *Saccharomyces, Saccharomycopsis (Yarrowia), Schizosaccharomyces; Neurospora, Podospora, Sordaria;* the sexual stages of both *Aspergillus* and *Penicillium; Ascobolus;* truffles and morels
Subdivision: Basidiomycotina ("club fungi"; septate mycelium or yeast; sexual spores borne exogenously on a basidium)
Examples: *Puccinia, Ustilago,* jelly fungi, rusts, smuts; *Agaricus, Coprinus, Schizophyllum,* mushrooms, puffballs, shelf fungi
Subdivision: Deuteromycotina (the Fungi Imperfecti; septate mycelium or yeast; no known sexual phase)
Examples: *Aspergillus, Fusarium, Penicillium; Candida, Histoplasma, Wangiella*

REFERENCES

Ainsworth, G. C. (1971). "Ainsworth & Bisby's Dictionary of the Fungi," 6th Ed. Commonw. Mycol. Inst., Kew, Surrey, England.
Ainsworth, G. C., Sparrow, F. K., and Sussman, A. S. (1973a). "The Fungi: An Advanced Treatise, Vol. 4A, A Taxonomic Review with Keys: Ascomycetes and Fungi Imperfecti." Academic Press, New York.
Ainsworth, G. C., Sparrow, F. K., and Sussman, A. S. (1973b). "The Fungi: An Advanced Treatise, Vol. 4B, A Taxonomic Review with Keys: Basidiomycetes and Lower Fungi." Academic Press, New York.

Alexopoulos, C. J., and Mims, C. W. (1979). "Introductory Mycology," 3rd Ed. Wiley, New York.

Bennett, J. W. (1985). Taxonomy of fungi and biology of the aspergilli. *In* "Biology of Industrial Microorganisms" (A. L. Demain and N. Soloman, eds.), pp. 359–406. Cummings, Menlo Park, California.

Burnett, J. H. (1968). "Fundamentals of Mycology." Arnold, London.

Burnett, J. H. (1975). "Mycogenetics: An Introduction to the General Genetics of Fungi." Wiley, New York.

Esser, K., and Kuenen, R. (1967). "Genetics of Fungi" (E. Steiner, transl.). Springer-Verlag, Berlin and New York.

Fincham, J. R. S., Day, P. R., and Radford, A. (1979). "Fungal Genetics," 4th Ed. Blackwell, Oxford.

Moore-Landecker, E. (1982). "Fundamentals of the Fungi," 2nd Ed. Prentice-Hall, Englewood Cliff, New Jersey.

Nelson, P. E., Toussoun, T. A., and Cook, R. J., eds. (1981). "Fusarium; Diseases, Biology, and Taxonomy." Pennsylvania State Univ. Press, University Park.

Pitt, J. I. (1979). "The Genus Penicillium and Its Teleomorphic States Eupenicillium and Talaromyces." Academic Press, New York.

Ramirez, C. (1982). "The Manual and Atlas of the Penicillia." Elsevier, Amsterdam.

Raper, K. B., and Fennell, D. I. (1965). "The Genus Aspergillus." Williams & Wilkins, Baltimore, Maryland.

Rippon, J. W. (1982). "Medical Mycology," 2nd Ed. Saunders, Philadelphia, Pennsylvania.

Ross, I. I. (1979). "Biology of the Fungi." McGraw-Hill, New York.

II

Conventions for Gene Symbols

J. W. BENNETT
Department of Biology
Tulane University
New Orleans, Louisiana

LINDA L. LASURE
Bioproduct Research
Miles Laboratories, Inc.
Elkhart, Indiana

I. *Aspergillus nidulans* 538
II. *Neurospora crassa* 539
III. *Saccharomyces cerevisiae* 540
IV. Other Fungi 542
References 542

Recommendations for uniform conventions of genetic nomenclature have been published for bacteria (Demerec *et al.*, 1966), *Aspergillus nidulans* (Clutterbuck, 1973), *Saccharomyces cerevisiae* (Sherman, 1981), and *Neurospora crassa* (Perkins *et al.*, 1982). In this volume, we make no attempt to impose a uniform standard of genetic symbols, but rather allow our authors to utilize the conventions of their particular organism and laboratory.

Although the designations for gene symbols and phenotypes are not the same for bacteria, yeasts, and molds, enough similarity exists to mislead the unwary reader. Since the publication of the proposals for bacterial genetics by Demerec *et al.* (1966), most primary gene symbols have been designated by three-letter, italicized symbols (e.g., *arg* for a locus affecting arginine biosynthesis). Some *Neurospora* and *Aspergillus* symbols predate the proposals for standardization of genetic nomenclature in bacteria and have fewer or more than three letters.

The conventions for distinguishing different loci that produce the same phe-

notypic change show minor, but confusing, variation from system to system. In bacteria and *A. nidulans* an italicized capital letter immediately follows the three-letter symbol (*argA*, *argB*, etc.), while in yeast nonhyphenated numbers are used (*arg1*, *arg2*, etc.). In *N. crassa*, hyphenated numbers are used to distinguish loci (*arg-1*, *arg-2*, etc.). In yeast, hyphenated numbers designate alleles (*arg1-37*); in *A. nidulans*, unhyphenated numbers designate alleles (*argA2*) and hyphenated numbers designate unmapped mutants (*arg-51*).

The conventions for phenotype, dominance, mating-type loci, designation of wild type, and other genetic symbols also show subtle differences between the systems. The most important recommendations from Clutterbuck (1973), Sherman (1981), and Perkins *et al.* (1982) are summarized in Sections I–III. Some representative examples are given to illustrate each system. Section IV cites a few additional systems of fungal genetic nomenclature. See the references for more complete explanations of all these nomenclatural conventions.

I. *ASPERGILLUS NIDULANS*

The recommendations for the nomenclature and conventions used for *A. nidulans* follow those of bacterial genetics and are published in Clutterbuck (1973). All genetic loci and mutants introduced subsequent to this publication are designated by three-letter symbols in italics (e.g., *arg*). Older symbols, previously adopted in the literature, are retained and consist of one to five italic letters (e.g., *y* = yellow; *panto* = pantothenic acid requirement). Nonallelic loci that have the same primary symbols are distinguished by an italic capital letter following the symbol, e.g., *argA*, *argB*. Alleles are distinguished by italic serial mutant numbers after the symbol and locus letter, e.g., *argA1*, *argA2*. Where the allelic relationships of a mutant have not yet been determined, the capital letter is replaced by a hyphenated number (e.g., *arg-51*). Mitochondrial gene symbols are enclosed in square brackets, e.g., [*oliA1*].

Wild-type alleles are indicated by a superscript "plus," e.g., *argA*$^+$. Occasionally, dominant mutants are designated by capitalizing the first of the three letters in a symbol (*Acr* for acriflavine resistance). In general, dominance is not indicated in the primary gene symbol. Symbols for phenotypes are distinguished from symbols for genes. Often the phenotype is simply written out in unabbreviated fashion (e.g., "arginine requirement"); alternatively, a nonitalic version of the gene symbol with the first letter capitalized is used, e.g., Arg$^-$.

Suppressors used to be designated by complex symbols including the locus and/or allele suppressed, e.g., *suA1adE20*, but now simple symbols are encouraged, e.g., *suaA1* allele-specific, locus-nonspecific suppressor. It is important to note that the wild-type, nonsuppressing allele is designated with a symbol "plus," as in *suaA*$^+$, opposite to the usage for bacteria.

Superscripts are used to indicate mutants with specific properties; for instance, *areA*^{d}*18* is an *areA* allele giving derepressed phenotypes for ammonium-repressed genes, while *areA*^{r}*1* gives correspondingly repressed phenotypes.

The following examples illustrate the conventions used in the genetic nomenclature for *A. nidulans*:

argA	A specific locus or mutation that produces a requirement for arginine as the phenotype
argA^{+}	The wild-type allele
argA2	A specific allele or mutation in the *argA* gene
arg-51	An arginine-requiring mutant not yet tested for allelism, whose locus is unknown
Arg^{+}	A strain not requiring arginine
Arg^{-}	A strain requiring arginine

A list of *A. nidulans* loci is given in Clutterbuck (1974), genetic maps are given in Clutterbuck (1984), and the mitochondrial genome is summarized by Spooner and Turner (1984).

II. *NEUROSPORA CRASSA*

A summary of conventions, gene symbols, and map locations of *N. crassa* genes is presented in Perkins *et al.* (1982), following Barratt and Perkins (1965). These conventions antedate bacterial genetic nomenclature and more closely follow those of *Drosophila*. Three-letter gene symbols are used most frequently, but symbols of one to four letters are also found. Two-letter symbols are quite common (e.g., *ad,* adenine requirement; *qa,* quinate utilization). Recessive gene symbols are written entirely in lowercase italics. When the mutant allele is known to be dominant, the first letter is capitalized (e.g., *Sk,* Spore killer).

Symbols without superscripts are used to represent mutant alleles. The same symbol with a superscript "plus" designates the wild-type allele, e.g., *ad*$^{+}$. Alleles differing in resistance or sensitivity, or allelic series having no definitive wild type, may be distinguished by other superscripts (e.g., *cyh-1*^{R}, cycloheximide resistance; *cyh-1*^{S}, cycloheximide sensitivity).

Nonallelic loci are distinguished from one another by numbers, separated from the symbol for the locus by a hyphen, e.g., *ad-1, ad-2*. The use of hyphens to distinguish nonallelic gene symbols differs sharply from the conventions for bacteria, *Aspergillus,* and yeast. In *Neurospora,* the allele number is "not usually displayed with the gene symbols, except when necessitated by the use of several alleles, when it is included in parentheses after the full locus symbol, e.g. *pyr-3* (KS43), or when a new mutant gene has not yet been assigned a locus number pending tests for allelism with similar genes at previously established

loci. In the latter situation, a mutant gene is temporarily designated by an appropriate letter symbol followed immediately by the allele number in parentheses, e.g. *ilv(STL6)*" (Perkins *et al.*, 1982, p. 427).

Mating-type alleles are called *A* and *a*. Suppressors are designated *su*, followed immediately by the symbol of the suppressed gene in parentheses; nonallelic suppressors of the same gene are distinguished by hyphenated numbers following the parentheses, e.g., *su(met-7)-1*, *su(met-7)-2*. Following the *Drosophila* convention, su^+ designates the wild type and *su* designates the mutant suppressor allele.

The following examples illustrate the major conventions used in the genetic nomenclature for *N. crassa:*

arg	Any locus or mutation that produces a requirement for arginine as the phenotype
arg-1	A specific locus that produces a requirement for arginine
arg-1$^+$	The wild-type allele of the *arg-1* gene
arg-1 (JWB7)	A specific allele of the *arg-1* gene
arg (JWB22)	An arginine-requiring mutant not yet tested for allelism, whose locus is unknown
Arg$^+$	A strain not requiring arginine
Arg$^-$	A strain requiring arginine

The Perkins *et al.* (1982) reference includes a detailed compendium of *N. crassa* loci and linkage maps. The maps are updated in Perkins (1984) and the mitochondrial genome is summarized by Collins and Lambowitz (1984).

III. *SACCHAROMYCES CEREVISIAE*

The recommendations for the nomenclature and conventions used in yeast genetics are summarized by Sherman (1981) and Sherman and Lawrence (1974). Gene symbols are consistent with the proposals of Demerec *et al.* (1966), whenever possible, and are designated by three italicized letters, e.g., *arg*. Contrary to the proposals of Demerec *et al.* (1966), the genetic locus is identified by a number (not a letter) following the gene symbol, e.g., *arg2*. Dominant alleles are denoted by using uppercase italics for all three letters of the gene symbol, e.g., *ARG2*. Lowercase letters symbolize the recessive allele, e.g., the auxotroph *arg2*. Wild-type genes are designated with a superscript "plus," (*sup6*$^+$ or *ARG2*$^+$). Alleles are designated by a number separated from the locus number by a hyphen, e.g., *arg2-14*. Locus numbers are consistent with the original assignments; however, allele numbers may be specific to a particular laboratory.

Phenotypic designations are written out or denoted by cognate symbols, with-

out italics, and by the superscripts "plus" and "minus." For example, independence of and requirement for arginine can be symbolized, respectively, as Arg^+ and Arg^-.

Gene clusters, complementation groups within a gene, or domains within a gene having different properties are designated by capital letters following the locus number, e.g., *his4A, his4B*. (Note that in the conventions of Demerec *et al.*, 1966, capital letters following the gene symbol designate different loci.)

Wild-type and mutant alleles of the mating-type and related loci do not follow the standard rules. The two wild-type alleles at the mating-type locus are designated *MAT***a** and *MAT*α. The two complementation groups of the *MAT*α locus are denoted *MAT*α*1* and *MAT*α*2*. Mutations of the *MAT* genes are denoted, e.g., *mat***a***-1*, *mat*α*1-1*. The wild-type homothallic alleles at the *HMR* and *HML* loci are denoted *HMR***a**, *HMR*α, *HML***a**, and *HML*α. Mutations at these loci are denoted, e.g., *hmra-1*, *hml*α*-1*.

The following examples illustrate the conventions used in the genetic nomenclature for *S. cerevisiae:*

ARG2	A locus or dominant allele
arg2	A locus or recessive allele that produces a requirement for arginine as the phenotype
ARG2$^+$	The wild-type allele of this gene
arg2–9	A specific allele or mutation at the *ARG2* locus
Arg^+	A strain not requiring arginine
Arg^-	A strain requiring arginine

For information on yeast mitochondrial genomes, see Grivell (1984).

For most structural genes that code for proteins, the nonmutant ("wild-type") allele is usually dominant to the mutant form of a gene. In yeast, the convention for dominant, "normal" genes utilizes capitalized italic symbols such as *HIS4* and *LEU2*. In traditional genetics, we learn about genes through their mutations, and linkage maps are created by following mutant alleles in crosses. Published linkage data, therefore, consist of gene symbols for the mutant, usually recessive, alleles [e.g., on linkage group III, *his4* and *leu2*. Those mutant alleles that are dominant to their nonmutant, "normal" alleles will appear on linkage maps in capital letters (*SUP22* and *FLD1* on chromosome IX)]. In addition, capital letters are used to represent dominant wild-type genes that control the same character and that are used for mapping (*SUC2, SUC1*, etc.), as well as DNA segments whose locations have been determined by a combination of recombinant DNA techniques and classical mapping procedures, e.g., *RDN1*, the segment encoding ribosomal RNA.

Detailed yeast linkage maps have been published by Mortimer and Schild (1980, 1984).

IV. OTHER FUNGI

Genetic conventions in other fungi sometimes follow one of the systems outlined above. In the past, workers with "less popular" species tended to follow some version of the bacterial–*A. nidulans* conventions; more recently, the yeast system has been gaining in popularity. For example, yeast conventions are used for the plant pathogen *Cochliobolus heterostrophus* (O. Yoder, personal communication). Regrettably, many workers adopt idiosyncratic symbols.

Both the "Handbook of Genetics, Vol. 1, Bacteria, Bacteriophages, and Fungi" (King, 1974) and "Genetic Maps 1984" (O'Brien, 1984) contain information about some of the better studied of the less popular fungi. Specific references, in alphabetic order by genus, follows:

Ascobolus immersus (Decaris *et al.*, 1974)
Dictyostelium discoideum (Newell, 1984)
Phycomyces (Cerdá-Olmedo, 1974)
Podospora anserina (Esser, 1974; Marcou *et al.*, 1984)
Schizosaccharomyces pombe (Gutz *et al.*, 1974)
Sordaria (Olive, 1974)
Ustilago maydis (Holliday, 1974)

Two species of Basidiomycetes, *Coprinus radiatus* and *Schizophyllum commune,* have been studied intensively, especially with respect to their incompatibility factors. Consult the following references for more information about these systems: Raper (1966), Guerdoux (1974), Raper and Hoffman (1974), and Schwalb and Miles (1978).

ACKNOWLEDGMENTS

We thank A. J. Clutterbuck, R. Metzenberg, D. Perkins, and F. Sherman for reviewing sections of this appendix.

REFERENCES

Barratt, R. W., and Perkins, D. D. (1965). Neurospora genetic Nomenclature. *Neurospora Newsl.* **8,** 23–24.

Cerdá-Olmedo, E. (1974). Phycomyces. *In* "Handbook of Genetics, Vol. 1, Bacteria, Bacteriophages, and Fungi" (R. C. King, ed.), pp. 343–357. Plenum, New York.

Clutterbuck, A. J. (1973). Gene symbols in *Aspergillus nidulans. Genet. Res.* **21,** 291–296.

Clutterbuck, A. J. (1974). *Aspergillus nidulans. In* "Handbook of Genetics, Vol. 1, Bacteria, Bacteriophages, and Fungi" (R. C. King, ed.), pp. 447–510. Plenum, New York.

Clutterbuck, A. J. (1984). Loci and linkage map of the filamentous fungus *Aspergillus nidulans*

(Eidam) Winder (n = 8). *In* "Genetic Maps 1984" (S. J. O'Brien, ed.), pp. 265–273. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Collins, R. A., and Lambowitz, A. M. (1984). The physical and genetic map of mitochrondrial DNA from *Neurospora crassa* strain 74-OR23-1*A*. *In* "Genetic Maps 1984" (S. J. O'Brien, ed.), pp. 274–276. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Decaris, B., Girard, J., and Leblon, G. (1974). Ascobolus. *In* "Handbook of Genetics, Vol. 1, Bacteria, Bacteriophages, and Fungi" (R. C. King, ed.), pp. 563–573. Plenum, New York.

Demerec, M., Adelberg, E. A., Clark, A. J., and Hartman, P. E. (1966). A proposal for a uniform nomenclature in bacterial genetics. *Genetics* **54,** 61–76.

Esser, K. (1974). *Podospora anserina. In* "Handbook of Genetics, Vol. 1, Bacteria, Bacteriophages, and Fungi" (R. C. King, ed.), pp. 531–551. Plenum, New York.

Grivell, L. A. (1984). Restriction and genetic maps of yeast mitochondrial DNA. *In* "Genetic Maps 1984" (S. J. O'Brien, ed.), pp. 234–247. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Guerdoux, J. L. (1974). Coprinus. *In* "Handbook of Genetics, Vol. 1, Bacteria, Bacteriophages, and Fungi" (R. C. King, ed.), pp. 627–636. Plenum, New York.

Gutz, H., Heslot, H., Leupold, U., and Oprieno, N. (1974). *Schizosaccharomyces pombe. In* "Handbook of Genetics, Vol. 1, Bacteria, Bacteriophages, and Fungi" (R. C. King, ed.), pp. 395–446. Plenum, New York.

Holliday, R. (1974). *Usilago maydis. In* "Handbook of Genetics, Vol. 1, Bacteria, Bacteriophages, and Fungi" (R. C. King, ed.), pp. 575–595. Plenum, New York.

King, R. C., ed. (1974). "Handbook of Genetics, Vol. 1, Bacteria, Bacteriophages and Fungi." Plenum, New York.

Marcou, D., Picard-Bennoun, and Simonet, J.-M. (1984). Genetic map of *Podospora anserina. In* "Genetic Maps 1984" (S. J. O'Brien, ed.), pp. 252–261. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Mortimer, R. K., and Schild, D. (1980). Genetic map of *Saccharomyces cerevisiae. Microbiol. Rev.* **44,** 519–571.

Mortimer, R. K., and Schild, D. (1984). Genetic map of *Saccharomyces cerevisiae. In* "Genetic Maps 1984" (S. J. O'Brien, ed.), pp. 224–233. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Newell, P. C. (1984). Genetic loci of the cellular slime mold *Dictyostelium discoideum. In* "Genetic Maps 1984" (S. J. O'Brien,ed.), pp. 248–251. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

O'Brien, S. J., ed. (1984). "Genetic Maps 1984. A Compilation of Linkage and Restriction Maps of Genetically Studied Organisms," Vol. 3. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Olive, L. S. (1974). Sordaria. *In* "Handbook of Genetics, Vol. 1, Bácteria, Bacteriophages, and Fungi" (R. C. King, ed.), pp. 553–562. Plenum, New York.

Perkins, D. D. (1984). *Neurospora crassa* genetic maps. *In* "Genetic Maps 1984" (S. J. O'Brien, ed.), pp. 277–285. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Perkins, D. D., Radford, A., Newmeyer, D., and Björkman, M. (1982). Chromosomal loci of *Neurospora crassa. Microbiol. Rev.* **46,** 426–570.

Raper, J. R. (1966). "Genetics of Sexuality in Higher Fungi." Ronald Press, New York.

Raper, J. R., and Hoffman, R. M. (1974). *Schizophyllum commune. In* "Handbook of Genetics, Vol. 1, Bacteria, Bacteriophages, and Fungi" (R. C. King, ed.), pp. 597–626. Plenum, New York.

Schwalb, M. N., and Miles, P. G., eds. (1978). "Genetics and Morphogenesis in the Basidiomycetes." Academic Press, New York.

Sherman, F. (1981). Genetic nomenclature. *In* "Molecular Biology of the Yeast, Saccharomyces

cerevisiae'' (J. N. Strathern, E. W. Jones, and J. R. Broach, eds.), pp. 639–640. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Sherman, F., and Lawrence, C. W. (1974). Saccharomyces. *In* ''Handbook of Genetics, Vol. 1, Bacteria, Bacteriophages, and Fungi'' (R. C. King, ed.), pp. 359–393. Plenum, New York.

Spooner, R. A., and Turner, G. (1984). Mitochondrial genetic loci of *Aspergillus nidulans*. *In* ''Genetic Maps 1984'' (S. J. O'Brien, ed.), pp. 262–264. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Index

A

Abortive transformants, 283
Absidia, 534
Accumulation of protein, 454
Acetamidase, 346
Acetate requirement, 323
Achlya, 534
Achlya ambisexualis, 70
Achlya bisexualis, 43
Actin, 484
  actin gene, 461
acuD, 273, 347
Additive integration, 169
Adenine nucleotide translocator, 88
Adenylosuccinase, 8
Aflatoxin, 49
Agaricus, 534
Alanine, 370
alcA alcR cluster, 350
Alcohol dehydrogenase I, 346
Alcohol utilization gene cluster, 346
alcR, 313, 330, 331, 336
Aldehyde dehydrogenase, 346
Allele-specific suppression, 321, 330
Allelic complementation, 8, 11
Allomyces arbuscula, 42, 49, 534
Allosteric regulation, 22
Alpha-aminoadipate, 129
Alpha-aminoadipate aminotransferase, 201
Alpha-factor, 215
Alpha-ketoglutaric acid, 368
am+ 248
Ambrosiozyma 54
amdS, 264, 266, 267, 539, 547, 555
Amino acid
  control, general, 179
  control, of transcription, 210
  sequence homology, 290
  uptake, 9
DL-α-Aminoadipic acid, 201
p-Aminobenzoic acid biosynthesis, 347
Aminoglycoside phosphotransferase, 199
Amphidiploids, 46
cAMP-binding proteins, 452
cAMP-dependent kinases, 452
Amplification, 198, 305
Amylases, 49, 496, 497
Amyloglucosidase, 497
Amylopectin, 496, 498
Amylose, 496, 497
Aneuploid, 46, 144, 298
Ans1, 271
Antibiotic yield, 294
Antibiotic-resistant transformants, 200
Antibiotics, 198, 293, 294, 305
Anticodon, 7, 8
aplA, 313, 316, 318, 319, 330, 331, 349
Apocytochrome *b*, 70, 80
L-Arabinose regulon, 315
D-Arabitol, 382
araC, 315
areA, 313, 316, 319, 320, 321, 322, 330, 334
areB, 322, 328
arg-12, 284, 285, 286, 290
Arginine metabolism, 282
Arginine permease gene, 217
aro, 11
ARS (autonomously replicating sequence), 136, 171, 172, 173, 177, 178, 198, 210, 211, 245, 271, 272, 435, 436
  Cochliobolus, 435
  mitochondrial, 411
Artificial chromosomes, 177
Ascobolus, 532
Ascobolus immersus, 540
Ascomycotina (Ascomycetes), 36, 53, 499, 531, 532, 534
Ascospores, 54
Ashbya gossypii, 378
Aspartic acid, 370, 371
Aspergillus, 49, 72, 82, 83, 87, 264, 265, 382, 497, 507, 510, 534
Aspergillus amstellodami, 84, 86, 272, 432
Aspergillus awamori, 498, 506, 507
Aspergillus flavus, 49
Aspergillus fonsecaens, 362
Aspergillus genes, 297

Aspergillus griseus, 368
Aspergillus itaconicus, 365
Aspergillus leporis, 49
Aspergillus microgynus, 49
Aspergillus nidulans, 43, 67, 68, 71, 72, 73, 74, 75, 281, 282, 283, 284, 286, 288, 290, 291, 295, 298, 299, 309, 310, 312, 313, 314, 322, 334, 350, 362, 518, 533, 537, 538
Aspergillus niger, 67, 72, 268, 350, 361, 362, 363, 365, 366, 368, 372, 497, 507, 522
Aspergillus orchraceus, 375
Aspergillus oryzae, 49, 51, 371, 376, 522
Aspergillus parasiticus, 49
Aspergillus sojae, 49
Aspergillus tamarii, 49, 51
Aspergillus terreus, 365, 375
Aspergillus ustus, 370
Aspergillus wentii, 51
ATP synthetase, 347
ATPase 6, 70, 72, 73, 75, 76, 78, 79, 86
ATPase 8, 71, 72, 75, 78
ATPase 9, 71, 75, 79, 82, 83, 88
Aureobasidium pullulans, 372, 373
Autogenous regulation, 315
Autolysis, 37
Autonomous plasmid, 167, 177, 183, 280
Autonomous replication, 198
Autonomously replicating plasmid (YRp31), 208
Auxotrophic mutants, 3, 4, 6, 127, 207, 478

B

Balanced interactions, 304
Balanced lethal genotypes, 298
Base pairing, 41
Basidiomycotina (Basidiomycetes), 36, 52, 53, 55, 532, 534
Benomyl, 10, 131
Beta-carotene, 380
Beta-galactosidase, 274
Beta-galactosidase gene, 179, 180, 219
Beta-glucosidase, 502, 506, 507
Biomass, 496
Biopolymers, 372
bla gene, of *E. coli*, 503
Blakeslea trispora, 380, 534
Blastocladiella emersonii, 49, 534
Brettanomyces abstines/B. custersii, 51
Brettanomyces anomalus/B. clausenii, 51
Brettanomyces galactosidase, 151
Brettanomyces tubulin, 10
Britten and Davidson, 325
Brute force cloning method, 350

C

cAMP-binding proteins, 452
cAMP-dependent kinases, 452
Candida, 47, 48, 534, 364, 381
Candida albicans, 379, 476
Candida blankii, 367
Candida brumptii, 367, 368
Candida cloacae, 375
Candida guilliermondii, 371
Candida humicola, 370
Candida hydrocarbofumarica, 367
Candida lipolytica (Yarrowia), 363, 364, 368, 376, 377, 379
Candida maltosa, 461
Candida pelliculosa, 369
Candida polymorpha, 382
Candida rugosa, 367, 368
Candida sp. no. 107, 375
Candida stellatoidea, 477, 483, 485, 486
Candida tropicalis, 370, 477, 478, 483, 485
Candida utilis, 367, 369, 371, 377, 378, 381
Candida zeylanoides, 364
CAN1, 128, 199, 218
Capsule, 471
Carbamoyl phosphate, 281
Carbamoyl-phosphate synthetase, 288
Carbohydrases, 494, 505, 508
Carbon catabolite repression, 22, 322, 323, 328, 329
Carbon sources, 323
Carotenoids, 379, 380
Cascade experiments, 349
Cascade, of regulatory genes, 22
CAT, 219
CAT cartridges, 219, 220
cat gene, 219
catabolic dehydroquinase, 248
CBHI, 505, 506
CBPI locus, 89
cdc6, 131
CDC8, 180

cDNA, 349
cDNA clone, 330
Cell
  cycle, 476
  morphology, 452
  size, 481
  surface antigens, 54
  wall, 455
  wall, mannans, 53, 54
  wall, matrix polymers, 456
Cell-cycle mutants, 10
Cellobiohydrolase (exo-β-1,4-glucan cellobiohydrolase, E.C. 3.2.191), 504
Cellodextrins, 504
Cellular localization, 151
Cellular senescence, 83
Cellulase, 494, 507, 509
Cellulose, 503
CEN, 177, 178
Centromere, 168, 174, 177, 437
Centromere plasmids, 136
Cephalosporium acremonium, 68, 523
Cephalosporium eichhorniae, 372
Ceratoycystis, 48
Cerulinin, 456
Cesium chloride gradient, 38
Chaetomium cellulyticum, 372, 375
Chain termination mutations, 11, 336
Charon 4a clone, 333
Chimeric plasmid DNA, 162
Chitin synthase, 455
Chloramphenicol acetylase, 265
Chloramphenicol acetyltransferase, 135, 219
Chromatin, 27, 337
Chromosome
  chromosomal aberration screening method, 347, 350
  chromosomal aberrations, 350
  chromosomal DNA, 207
  chromosomal rearrangements, 152, 322, 328
  duplication, 295
  ends (telemores), 177
  exchanges, 185
  loss, 147
  loss mapping, 130
  mini-chromosome, 175
  replicators, 172
  sequences, 165
  stability, 177
Chytridiomycetes, 52, 53
cis-acting
  receptor site, 325
  regulatory elements, 309, 315, 321, 322, 327, 328, 329, 330, 333
cis effects of mutations, 12, 22, 23, 24, 179
Citrate synthase, 90
Citric acid, 361, 365
Citrulline, 281
Cladosporium, 375
Claviceps purpurea, 68, 83, 84, 369
Clone, 330, 333, 336
Cloning
  by complementation, 198
  gene, 141
  methodologies, 310, 345, 352
  S. cerevisiae mating type, 26
  strategy, 345
  vectors, 409
Cluster genes, 10, 12, 14, 16, 332, 346
Cochliobolus heterostrophus, 69, 83, 84, 432, 542
Codominant, 317
Codon usage, 27, 234, 501, 507
COI, COII and COII, 72, 73, 74, 75, 78, 80, 87, 89, 90
Coinducer, 327
Cold-sensitive mutants, 10
Colinearity, 6
Colony hybridization, 165
Complementation, 9, 25, 34, 143, 152, 179, 198, 208, 281, 311, 315, 319, 333, 346, 461, 478, 481, 485
Consensus, 503
Conspecific, 51
Constitutive, 316, 317
Constitutive mutations, 13, 313
Construction
  of genetic map, 129
  of library, 142
Control of gene expression, 309
Copper-binding protein, 200
Coprinus, 534
Coprinus cinereus, 67, 69
Coprinus radiatus, 542
Coprinus stercorarius, 69
Copy number, 297, 305
  control, 137
  effects, 144
Correction-deficient mutants (cor), 183
Cosmid, 274

Cosmid clones, 350
Cot, 44
Cotransformation, 267
cpc, 288
creA, 322
Cross-pathway regulation, 288
Cryptococcosis, 471
Cryptococcus laurentii, 370, 472
Cryptococcus laurentii var. *flavescens*, 55
Cryptococcus laurentii var. *magnus*, 55
Cryptococcus laurentii var. *laurentii*, 55
Cryptococcus neoformans, 471
Cryptococcus terricolus, 375
CUP1 (copper resistance), 185, 200
Cutinase, 421
Cyclic 3′,5′-adenosine monophosphate (cAMP), 452
Cycloheximide, 10
CYC1, 219
CYC1-*lacZ* fusions, 179
CYH2, 199, 218
Cytochrome
  cytochrome *b*, 72, 73, 75, 76, 78, 79, 80, 81, 89, 90
  cytochrome *bc*, 88
  cytochrome *c*, 6, 7, 25, 88
  cytochrome *c* oxidase, 83, 88
  cytochrome *c* oxidase subunits, 70, 72, 73, 74, 75, 78, 80, 87, 89, 90
  cytochrome *c* peroxidase, 88
  cytochrome oxidase, 289
Cytoduction, 137
Cytoplasmic gene expression, 281
Cytoplasmic mutants, 407
Cytosolic marker enzyme, 289

D

D-Arabitol, 382
DAPI, 43
Debaryomyces, 381
Debaryomyces condertii, 368
Debaryomyces marama, 39
Debaryomyces vanriji, 375
Dekkera bruxellensis/Brettanomyces lambicus, 51
Deletion, 317, 329, 332, 333, 334
Deletion mapping, 311
Delta–delta recombination, 186
Dematiaceous fungi, 474
2-Deoxygalactose, 199
Deoxyribose nucleic acid, *see* DNA
Derepression, 288
Detection, of gene, 460
Deuteromycotina (Deuteromycetes), 36, 503, 532, 534
Developmentally regulated gene clusters, 347
Dicistronic mRNA, 333, 335
Dictyostelium, 518, 534
Dictyostelium discoideum, 542
Diethyl pyrocarbonate, 38
Differential hybridization, 346
Differential protein synthesis, 457
Dimorphism, 450
p-Diphenoloxidase, 347
Diploids, 198, 207, 311, 319
Diploid, partial, 295, 296
Direct selection, 127
Direction, of transcription, 332
Disease, 469
Disomic, 298, 299, 300, 302, 303, 304, 305
Disomic strain stability, 295, 298
DL-α-Aminoadipic acid, 201
DNA
  base composition, 35, 38, 39
  complementarity, 45
  conformation, 28
  dispersed repetetive, 462
  divergence, 46
  fungal, 37
  genomic library, 205
  hetero-duplex, 182, 183, 184
  homology, 46, 54
  isolation, 70
  ligase, 10
  linear, 42, 176
  mitochondrial, 39, 176
  preparation for cloning, 142
  purity, 38
  reassociation, 44
  relatedness, 41, 44, 149, 183, 184
  renatured, 41, 44
  repair, 41
  replication origins, 171
  sequence analysis, 210
  telomeric, 177
  transforming, 165

uptake, 163
uptake by yeast cells, 162
DNA–DNA hybridization, 290
DNA–DNA reassociation, 46
DNases, 38
Dolipores, 52, 54
Domains, 314, 316
Dominance, 311, 317
Dominant markers, 134
Dose effects, 317, 319
Double-strand gap, 148, 183, 184
Down-promoter alleles, 329
Drug resistance, 9, 264
Duplication, 149
of genes, 129
of *penB2* locus, 299

E

Echinocandin B, 128
EF-1α, 458, 462
Effector, 317, 323
Elongation factor Tu, 88
Emericella nidulans, 533
Endo I, 505
Endo II, 506
Endoglucanases (endo-β-1,4-glucan glucanohydrolase E.C. 3.2.1.4), 503, 505, 508
Endomycopsis vernalis, 375
Endosymbiont hypothesis, 81
Enhancer-like element, 333
Enolase promoter, 506
ENO1, 200
Enrichment, 127, 128
Enzyme correlates of morphogenesis, 452
Epidermophyton, 41
Episomal, 168
Eremothecium ashbyii, 378
Ergosterol, 380
Erythorbic acid, 368
Escherichia coli, 198, 296, 315
Ethanol, 382, 383
supplementation, 323
utilization, 331
Ethidium bromide, 43
Eukaryotic cloning system, 404
Eumycota, 517, 532, 534
Evolutionary clock, 52
Evolutionary relationships, 55
Excision, 186
Exo-β-glucanases, 55
Exon, 182
Exon/intron junctions, 499
Expression, 1, 151, 180

F

Fatty acid, 375, 376, 456
Fatty acid-less death, 128
Fiji plasmid, 246
Filamentous fungi, 280, 281, 291, 345
Filobasidiella bacillispora, 46
Filobasidiella neoformans, 46, 471
Fine-structure mapping, 311, 316
Flanking markers, 311
Flavor nucleotides, 377
Flow cytometry, 352
Fluoroacetate, 152
5-Fluorocytosine, 482, 484
5-Fluoro-orotic acid, 199
Flux, 106
Fragment patterns, 49
Frameshift mutations, 25
Fumaric acid, 366, 367
Fungi
associations of, 521
carbohydrases of, 493
Fungi Imperfecti, 533
of industrial importance, 352
plasmids of, 297
taxonomy of, 531
Fusarium, 371, 503, 534
Fusarium graminearum, 371
Fusion protein, 151
F1 progeny, 46
F1-ATPase, 88, 90

G

G + C content, 38
Galactose permease-negative (*gal2*), 199
Galactose-negative mutants, 199
gal mutants
gal1, 199
gal4, 199
GAL4c, 199
gal80, 199

Gapped
molecules, 183
plasmid, 210
GDH, 454
Gelasinospora austosteira, 48
Gene
amplification, 185, 296, 297
bank, 273
cluster, 10, 12, 18, 330, 332, 334, 349
conversion, 169, 182, 183, 185, 213
disruption, 145, 179, 213, 217
dosage, 296, 297
dosage effects, 145, 179
dosage on penicillin production, 298
duplication, 296, 305
essential, 146
eviction, 148, 179
expression, 151, 180
fine-structure mapping, 6
function, 53
fusion, 151, 179
heterologous, 182, 281, 291
homologous, 290
isolation, 178
isolation by self-cloning, 273
order, 74, 75
regulation, 13
replacement, 210, 212, 213
transplacement, 179, 210, 211
tRNA, 71, 74
Genetic
backgrounds, 301
code, 71, 74
divergence, 36
hybridization, 35, 44
map, 127, 147, 207
recombination, 182, 184
relatedness, 45
Genome size, 43, 66
Genomic "libraries," 25
Geotrichum, 41
Glucoamylase, 495, 497, 507, 508, 509
Gluconic acid, 365, 366
Glucose catabolic pathways, 453
Glucose-6-phosphate dehydrogenase, 289
Glusulase, 37, 163, 237
Glutamate dehydrogenase, 6, 7, 8, 25, 250, 321, 453
Glutamic acid, 370
Glutamine synthetase, 55, 321
L-Glutamine, 321
Glycerol, 380, 381
Glycogen, 497
Glycoprotein, 495, 504, 508
Glycosidases, 508
5′-GMP, 377
Gratuitous inducers, 315
Growth habits, 312
Growth testing, 312
Guanylic acid, 376
G418, 135, 199, 200, 265
G418-resistance gene, 200, 219

H

Hanseniospora valbyensis, 370
Hansenula, 53, 364, 375, 381
Hansenula alni, 46
Hansenula americana, 46
Hansenula anomala, 369
Hansenula beckii, 54
Hansenula bimundalis, 46, 53, 54
Hansenula canadensis, 48, 53, 54
Hansenula jadinii, 47
Hansenula minuta, 47, 48
Hansenula mrakii, 66, 69
Hansenula petersonii, 69
Hansenula polymorpha, 369
Hansenula wingei, 42, 46, 48, 53, 54
Haploid genome, 127
Helicobasidium mompa, 365
Hetero-duplex, 182, 183, 184
Heterokaryon, 138, 152, 311, 319, 362, 478, 484
Heterokaryon formation, 460
Heterologous expression, 345
Heterologous genes, 281, 291
Heterothallic strains, 45
Hierachy of structural genes, 321
High-copy-number vectors, 208
his genes
his-3, 11
HIS4, 179, 186
his7, 207
Histoplasma, 534
Histoplasma capsulatum, 473
Holliday junction, 183, 184
Homologous genes, 208, 290
Homologous recombination, 147, 205, 211, 266
Homologous transformation, 286

Homothallic, 45, 310
Humicola, 39
hxnR, 330, 331
hxnS, 349
hxyR, 349
Hybrid
  arrest of translation, 147
  enzyme, 9
  shuttle vector, 503
Hybrid-selected RNA, 147
Hybridization, 44
  with isologous gene, 346
Hydroxylapatite fractionation, 38

I

Immunochemical comparisons, 52
Immunoprecipitation, 289
5′-IMP, 377
Induction, 325
Industrial fermentations, 36, 357
Infection structures, 418
Initiation site, 77
Initiator, 330, 335
  constitutivity, 330
  sites, 314
Inosinic acid, 376
Inositol starvation, 128
Inserted gene sequence, 36
Insertion, 317, 350
Insertional translocation, 334, 335
intA, 325, 327, 329
Integration, 65, 168, 169, 170, 179, 205, 265, 267, 268, 286, 310, 333, 334
Integrative plasmid (YIp600), 208
Integrative transformants, 216
Integrative vectors, 135
Integrator genes, 309, 313, 325
Intervening sequence, 232, 498
Intracellular cAMP, 454
Intron, 26, 27, 72, 73, 80, 84, 182, 503
  group I, 80, 84
  group II, 80
  intron/exon junctions, 503, 507
  splicing, 79, 89
Inversion, 317, 321, 322, 327, 350
Invertase, 219
Inverted repeats, 70
Ionic strength, 43
Isoascorbic acid, 368
Isocitrate lyase, 273, 347
Isolation of mutants, 127, 459
Issatchenkia occidentalis, 46, 47
Issatchenkia orientalis, 46, 47
Issatchenkia scutulata, 45, 46
Issatchenkia scutulata exigua, 45
Issatchenkia scutulata var. *scutulata*, 46
Itaconic acid, 365

J

Jelly fungi, 534

K

kar1, 138
Kinetic complexity, 43
Kinetic modeling, 106
Kinetics, of duplex formation, 42
Kloeckera africana, 69
Kluyveromyces dobzhanskii, 45
Kluyveromyces fragilis, 45, 371, 379, 506, 508
Kluyveromyces lactis, 78
Kluyveromyces marxianus, 45
Kluyveromyces thermotolerans, 45
Koji, 49

L

L-Arabinose regulon, 315
L-Glutamine, 321
L-rRNA, *see* Ribosomal RNA
Labelle plasmid, 246, 410
Lactic acid, 367
lacZ, 151, 179, 180, 219
Lambda clones, 350
Leader sequences, 508
Leaky growth, 312
leuB gene, 476, 485
LEU2 gene, 213
Libraries, 222
Ligninases, 509
Linear DNA, *see* DNA
Linear plasmids, *see* Plasmids
Linearization (gap formation), 183
Linkage, 146, 165, 207, 472, 482
Lipid, 374, 376
Lipid biosynthesis, 456
Lipomyces lipofer, 375
Lipomyces starkeyi, 375
Liposome-mediated transformation, 280
Loss-of-function mutations, 313, 327

Lysine, 369, 370
Lysine biosynthesis, 203, 204
Lysozyme, 505
lys genes
  LYS2, 129, 198, 200, 203, 205, 206, 207, 208, 209, 211, 214, 215, 216, 217, 220, 222, 223
  lys2, 201
  LYS2+, 201
  LYS5, 129, 201, 203

M

Malic acid, 367
Maltodextrin, 497, 508
Mannans, 374, 455
Mannitol, 382
Marker, 211, 213
Marker rescue, 221, 269, 270
Mastigomycotina, 36, 532, 534
MAT, 215, 216, 217
Mating, 36, 46
Mating types, 459
Maturases, 72, 74, 80, 89
Mauriceville plasmid, 246
Meiotic stability of transformants, 285
Melanin, 471, 475
Meselson-Radding model, 183, 184
Metabolic compartments, 105
Metabolic versatility, 309, 311
Metabolism *in vivo,* 106
Methionine, 370
Methylation of EF-1α, 458, 459
Methylbenzimidazole-2-yl carbamate (MBC), 131
MET15, 128
MFα1, 215
MFα1 promoter, 220
MFα1–LYS2, 221
MFα1–SUC2 fusion gene, 217
Microsporum, 41
Mini-chromosomes, 175
Miso, 49
Mitochondria, 39, 176, 282, 291
  ATP synthetase, 264
  biogenesis, 281
  DNA (mtDNA), 271, 272, 410, 474, 484, 485
  DNA polymorphisms, 51
  DNA relatedness, 35, 49
  DNA restriction patterns, 50
  function, 90, 454
  gene expression, 281
  genome, 50, 66
  membrane marker, 289
  plasmids, 246, 407
  proteins, 290
Mitosis
  mitotic instability, 251
  mitotic stability of transformants, 285
Mitotic and meiotic pairing, 152
Modeling metabolism, 105
Molecular analysis, 310
Monomorphic mutants, 454
Monosomic, 207
Morels, 534
Mortierella, 534
Mortierella vinacea, 375
mRNA, 25, 216, 349
MSS51 locus, 89
Mucor, 371, 534
Mucor bacilliformis, 454
Mucor circinelloides, 375
Mucor genevensis, 376, 452, 456
Mucor mucedo, 459
Mucor racemosus, 449
Mucor ramannianus, 376
Mucor rouxii, 449
Multicellular form, 474
Multifunctional gene products, 10
Multiple-domain, 11
Mutagenesis, *in vitro*, 150, 179, 184
Mutagenesis, and screening, 359
Mutanase, 37
Mutants, *see* specific gene name
Mutation and selection, 294
Mutations, 294, 303, *see also* specific gene name
  increasing penicillin yield, 295
  reducing penicillin yield, 294, 295
Mutator locus, 73
Mycosis, 469
Myxomycota (myxomycetes), 52, 517, 532, 534

N

N-terminal presequence, 90
NADP-linked glutamate dehydrogenase, 346
Nalidixic acid, 482
NAM2 locus, 89
Nectria haematococca, 422

Neoclassical genetics, 127
Neurospora, 4, 43, 48, 49, 69, 71, 74, 82, 83, 85, 88, 245, 264, 281, 283, 286, 288, 289, 290, 291, 321, 508, 534
Neurospora crassa, 11, 48, 67, 72, 73, 75, 80, 83, 84, 85, 229, 280, 281, 283, 284, 286, 288, 289, 290, 291, 310, 321, 431, 499, 518, 533, 537, 539
Neurospora crassa, LA14 (*arg-12*), 283
Neurospora dodgei, 48
Neurospora intermedia, 83, 84
Neurospora qa cluster, 12
Neurospora transformation, 248
nib mutation, 69
Nick translation, 41
Nicotinate utilization, 316, 331
Nicotinic acid, 379
nirA, 314, 315, 318, 319, 331, 334
nit-2, 321
Nitrate assimilation, 309, 311, 314, 331, 332, 334
Nitrate reductase, 315, 320, 334
Nitrocellulose filters, 41
Nitrogen
  metabolite repression, 319, 321, 322, 328, 329, 334
  nitrogen-catabolite repression, 22
  nitrogen-starved growth, 312
  sources, 201, 312, 319, 322, 323, 328
Noncoding sequences, 67
Nonhierarchical heterogeneity, 321, 327
Nontandem duplications in yeast, 185
Northern blot, 333, 336
Novozyme 234, 260
Nuclear
  nuclear division in *Aspergillus nidulans*, 298
  nuclear genes, 88
  nuclear limitation, 317
  nuclear–cytoplasmic interactions, 88, 520
Nucleotide
  sequence homology, 290
  substitution, 52
Nutritional screening, 312
Nystatin, 482

O

ODC, 454
oliC, 265, 347
Oligomeric plasmid, 252
Oligomycin resistance, 264, 265
Oligonucleotide cataloging, 52
One gene–one enzyme concept, 5
One gene–one polypeptide concept, 7
One gene–one RNA transcript concept, 7
One integrator gene–one context, 325
One-step gene disruption, 150, 214, 215, 216
One-step gene replacement, 217, 218
Oomycetes, 39
Operon, 330
ORFs, 73, 74, 80, 89
Organic acids, 360
Origins of replication, 173, 177, 208
Ornithine, 281
Ornithine carbamoyltransferase, 281, 282, 286, 288, 289, 290
Ornithine carbamoyltransferase (*arg12^s*), 283
Ornithine decarboxylase, 454
Outer membrane polypeptide, 88
Overproducers, 295

P

p-Aminobenzoic acid biosynthesis, 347
p-Diphenoloxidase, 347
pabaA, 347
pACYC184, 500
Paecilomyces varioti, 367
palcA, 323, 324
pALS1, 249
pALS2, 249
Parasexual, 294
  parasexual cycle, 269, 310
Partial digest, 142
Partial loss-of-function mutations, 321
Pathogenesis, 418, 439, 520
Pathway-specific regulation, 309, 313, 315, 324
pBB116, 282, 283
pBR322, 207, 209, 283, 503
pBTI-I cosmid, 507
pDA6200, 205, 206, 207, 208, 209, 210
pDP2, 283, 285, 286, 288
Pellicularia, 376
pen, 298
penA1, 297, 299, 300
penB2, 297, 298, 299, 300, 301, 302, 304, 305
penB2$^+$, 300, 302, 304
penC3, 297
Penicillia, 350

Penicillin, 295
  titer (yield), 294, 295, 303, 304
Penicillinase, 296
Penicillium, 503, 534
Penicillium chrysogenum, 69, 294, 295, 297, 365, 371
Penicillium citrinum, 377
Penicillium cyclopium, 371
Penicillium glaucum, 366
Penicillium lilacinum, 375
Penicillium notatum, 366, 368
Penicillium purpurogenum, 365
Penicillium varioti, 371
Permeases, 324
pet$^-$ mutants, 89
Petite mutants, 78, 80, 81, 84, 272
PET484 locus, 90
pH regulation, 324, 325
Phaeohyphomycosis, 474
Phaeomycotic cysts, 474
Phanerochaete, 503, 505
Phanerochaete chrysosporium, 509
Phenylalanine, 370, 371
Phosphoglycerate kinase, 347
Phosphohexans, 374
Phosphorus repression, 323
Phycomyces, 542
Phycomyces blakesleeanus, 42, 52, 68, 380
Phylogenetic relatedness, 52, 55
Phylogeny, 35
Physarium, 518
Physarum polycephalum, 66
Physiological inducer, 315
Phythium, 534
Phytophthora, 534
Phytophthora infestans, 68
Pichia, 53, 54, 55, 364, 381
Pichia amylophila, 45, 46
Pichia haplophila, 382
Pichia lindneri, 47, 48
Pichia mississippiensis, 46
Pichia ohmeri, 54
Pichia quercibus, 382
Pilobolus, 534
Pisatin demethylase, 422
pJR2, 250
Plasma membrane function, 456
Plasmid, 42, 83, 84, 86, 87, 141, 200, 222, 240, 245, 270, *see also* pB; YEp
  excision, 210
  integration, 210
  integration targeting, 141
  "killer," 408
  Labelle, 246, 410
  linearization, 137, 184, 205, 211
  Mauriceville, 246
  mitochondrial, 246, 407
  native, 430
  nuclear, 408
  plasmid-like DNAs (PLDs), 246
  plasmid-like 2-μm circular DNA, 42, 200
  segregation, 172
  vectors, 197, 198, 200, 208
  YE, 211
  YR, 211
  YC, 211
  YI, 211
Plasmodiophorales, 53
Pleiotropic, 300
Podospora, 81, 83, 87, 534
Podospora anserina, 10, 69, 75, 83, 84, 273, 430, 542
Point mutations, 55
poky, 84, 85
Polar effects of mutations, 11
Polyadenylated RNA, 336
Polyamines, 281, 453
Polyethylene glycol (PEG), 163, 262
Polyfunctional polypeptides, 332
Polyol, 380, 381, 382
Polypeptide chain-termination, 317, 320, 321
Polyploids, 362
Polysaccharides, 372
Positive-acting regulatory gene *nir*, 313
Posttranslational modifications, 316, 457, 504
Precursor, 498
Primary metabolism, 357
Primary transcript, 76, 77, 78
prn gene
  prnA, 313, 329, 331, 332, 333
  prnADBC, 347, 349
  prnADBC cluster, 351
prnC, 267
Processing site, 79
Product accumulation, 105
Proline
  auxotrophy, 129
  catabolism, 309, 311, 332
  toxicity, 332
  utilization, 312, 331, 347

Promoter, 76, 78, 79, 235, 274, 321, 335, 498, 503, 507
fusion, 152, 218
qa, 235
region, 505
Protein
methylation, 458
polymorphism, 55
synthesis, 455
Proton magnetic resonance (PMR), 53
Protoplast, 36, 260, 472, 477, 478, *see also* Spheroplast
fusion, 364, 365
regeneration, 163
Pseudoconstitutive mutations, 316, 318, 331
Pseudogene, 81, 322
*Pst*I palindromes, 247
Puccinia, 534
Pullulan, 372, 373
Pulsed field gradient electrophoresis, 352
Purine degradation, 312, 315, 331
pVK55, 248
pyr gene
pyr-3, 11
pyr-44, 264, 267, 269, 271
pyrF, 476, 485
pyrG, 264, 267
Pyridoxine, 379
Pyroglutamyl residue, 505
Pyruvate dehydrogenase, 323
Pyruvic acid, 368
Pythium, 375
Pythium ultimum, 68

Q

qa gene cluster, 229, 233, 234
qa-2^{+}, 248
Quinate utilization, 347
Quinic acid, 230
qutBCE, 347

R

rad52, 131, 183, 184, 185
Ragged mutants, 84, 86
Rate-limiting steps, 105
Rational approaches for strain development, 305
recA, 270
Receptor sites, 309, 317, 321, 327, 328, 330
Recombinant DNA (rDNA), 36, 360, 384, 385, 460, 519, 541
Recombinant DNA methodology, 223
Regeneration, 480
Regulation of gene expression, 13, 309, 310, 321, 327
Regulatory circuit, 325
Regulatory gene, 309, 312, 313, 318, 322, 325
Regulatory gene *alcR*, 317
Repair, 149, 183, 184
Repeat sequences, 43, 67
Repetitive DNA, 325, 474
Replicating vectors, 270
Replication origins, 77, 78, 271
Reproduction and sexuality, 523
Restriction endonucleases, 49, 50, 210
Restriction pattern analysis, 51
Reverse transcription, 25
Rhizopus, 368, 371, 375, 497, 534
Rhizopus arrhizus, 366, 367
Rhizopus japonicus, 366
Rhizopus nigricans, 366
Rhizopus oryzae, 367
Rhodotorula, 371, 472
Rhodotorula glutinis, 371
Rhodotorula gracilis 375
Riboflavin, 378, 379
Ribosomal DNA, 474
Ribosomal proteins, 10, 52, 457
Ribosomal repeat, 267
Ribosomal RNA, 71, 72, 73, 75, 80, 87, 89
Ribosomal RNA relatedness, 51
Ribosome
subunit protein, S6, 457
Ribosome-binding site, 321
RNA, 38, 85
5 *S*, 52
9 *S*, 78, 79
16 *S*, 51
18 *S*, 52
blot analysis, 333
polymerase, 77
ribosomal, 76, 77, 78, 79, 86, 89, *see also* Ribosomal RNA
L-rRNA, 71, 72, 73, 75, 80, 86, 87, 89
S-rRNA, 71, 74, 75, 85
transfer, 7, 71, 74, 77, 78, 86
Rusts, 534

S

S nuclease, 42
S-adenosylmethionine, 458
Saccharomyces, 52, 54, 508, 534
Saccharomyces, industrial strains, 222
Saccharomyces bailii, 381
Saccharomyces carlsbergensis, 72
Saccharomyces cerevisiae, 6, 10, 42, 43, 67, 69, 71, 72, 80, 84, 87, 126, 197, 198, 199, 205, 206, 208, 223, 281, 282, 297, 310, 369, 370, 374, 379, 381, 383, 384, 506, 518, 533, 537, 540, 555
Saccaromyces diastaticus, 503
Sacchdromyces exiguus, 67, 69, 75
Saccharomyces microsporus, 379
Saccharomyces pombe, 72, 73, 75
Saccharomyces rouxii, 381
Saccharomyces telluris, 47
Saccharomycopsis, 534
Saccharomycopsis lipolytica (Yarrowia), 54, 69, 363, 364, 365, 369, 370
Saprolegnia, 41, 68, 534
Saprolegnia ferax, 49, 68
Scanning electron microscopy 54
Schizophyllum, 534
Schizophyllum commune, 69, 367, 368, 505, 509, 542
Schizosaccharomyces, 534
Schizosaccharomyces pombe, 7, 66, 69, 143, 542
Schwanniomyces, 47, 54
Schwanniomyces castellii, 48
Schwanniomyces occidentalis, 48
Scleroglucan, 373
Sclerotinia, 373
Sclerotium glucanicum, 373
Sclerotium rolfsii, 373
Scytalidium acidophilum, 372
Secondary metabolism, 358
Secondary metabolites, 293, 294, 296, 305, 522
Secreted enzymes, 324
Selectable markers (genes), 134, 198, 200, 211, 213, 264, 434
Selection of specific mutations, 128
Self-splicing, 80
Selfing, 310, 334
Semidominance, 317, 318
senDNAs, 87
Senescence, 84, 87, 88, 430
Septum ultrastructure, 52, 54
Sequence divergence, 51
Sequences, 503
Sexual cycle, 269
Sexual stages, 44
Shotgun 178, 200
Shuttle vectors, 10, 26, 142, 209, 253
  construction of, 410, *see also* Two-micron (2-μm) circle
Signal peptide, 505
Single-cell protein, 371
Single-copy nuclear DNA, 42
Single-stranded circular (ssc) DNA, 176
Site-specific recombination gene (FLP), 171
Slime mold, 350, 532
Smuts, 534
sod (stabilization of disomy) mutants, 298
$sod^{III}A$, 298, 299, 300, 302, 304
$sod^{III}A^{+}$, 304
$sod^{III}A1$, 304
Sordaria, 534
Southern blot hybridization, 166, 207, 211, 221, 284, 286, 329, 350
Species, 35, 36, 46
Sphacelotheca sorghii, 370
Spheroplast, 37, 133, 163, 280, 460, 472, 475
  fusion, 460, 472, 475, 477, 478
  transformation, 163
Splicing system, 182
spm (SPT), 186
Sporangiospores, 458
Sporobolomyces odorus, 371
Sporothrix schenckii, 48
Starch, 496
Steady-state model, 105, 106
Sterol, 456
Stopper mutant, 67, 84, 85, 86
Strain G96, 300
Strain improvement, 293, 294
Stromatinia 373
Structural gene, 312, 315, 325, 327, 333
suAmeth, 324, 329
Subunit 9, 347
Succinic acid, 368
SUC2, 217, 219, 220
Sugar metabolism, 154
Suillus grisellus, 66, 69
Sulfur repression, 324

Superoxide dismutase, 55
Superstrains, 130
  fungal species, 429
Suppression, 7
SUP11, 217
SUP4, 217
Surface structure, 54
Surrogate genetics, 153
Systems analysis, 106

T

tamA, 322
Tandem duplications, 211, 290
Tandem repeat, 268
Tartaric acid, 368
Taxonomy, 516
tcm1 (trichodermin resistance), 199
Telomeres, 137, 177
Telomeric DNA, 177, 178
Temperature-conditional mutants, 5, 10
Termination regions, 503, 506
Terminator, 503, 507
Tetrads, 183, 198, 207
  analysis, 129, 165, 216, 286
  unordered, 286
Tetrahymena, 80
Tetrahymena rDNA, 137
Tetrahymena thermophila, 52
Threonine, 370, 371
Thymidine kinase gene, 180
Tn601, 135, 199, 219
Tn9, 219
Tn903 (Tn55), 199
Torula, 371
Torulaspora, 52, 54
Torulopsis, 47, 66, 74, 364, 371
Torulopsis bovina, 47
Torulopsis candida, 375
Torulopsis glabrata, 42, 69, 75
Torulopsis magnoliae, 381
Torulopsis pintolopessi, 47, 48
Toxins, host-specific, 425
trans-Acting genes, 13, 19, 20, 313
Transcript processing, 78
Transcription, 77, 335
Transcriptional regulation, 208, 216
Transformation, 133, 201, 236, 260, 262, 280, 310, 317, 319, 330, 333, 334, 345, 351, 463
  abortive, 266, 271
  Aspergillus nidulans, 347
  efficiency, 162, 266, 271
  factors in *Neurospora*, 239
  high frequency, 135, 352
  integrative, 183
  intergeneric, 241
  lithium acetate, 238
  of nutritional mutants, 264
  Podospora, 409
  procedures 263
  qa and *Neurospora* 236
  types of *Neurospora* transformants, 240
  whole cell, 166
  yeast, 161, 162, 167, 178, 197, 198
Transition model, 105
Translational suppressors, 332
Translocation, 299, 311, 317, 320, 322, 335, 350
Transmission electron microscopy, 54
Transposable element, 81, 186, 328
Transposition, 83
Transposons, 199
Tremella mesenterica, 373
Trichoderma, 503, 504, 506, 510
Trichoderma cellobiohydrolase I, 508
Trichoderma reesei, 506
Trichophyton, 41
Trichosporon cutaneum, 370
Trigomopsis, 381
Trisomic, 298
Trisporic acids, 459
tRNA, 7, 71, 74, 77, 78, 84
tRNA genes, 461
tRNA processing locus, 79
trpC, 264, 266, 269, 274
TRP1, 207, 211, 312
TRP1 RI circle, 174
trp1-289, 211
Truffles, 534
Tryptic peptides of EF-1α, 459
Tryptophan, 369
Tryptophan synthetase, 6, 7
Tubulin, 10
Tubulin genes, 347
Two-micron (2-μm) circle, 26, 136, 169, 202, 245
Two-micron (2-μm) DNA, 207
Two-micron (2-μm) DNA plasmid, 198

Two-micron (2-μm) vector, 200
Ty, 169, 186, 222, 248

U

uaY, 315, 316, 317, 318, 319, 330, 331
uaZ, 349
Unequal sister chromatid exchanges, 185
Unique sequences, 42
Up-promoter effect, 330
Up-promoter mutation, 316, 329
Ura$^+$*Lys*$^+$ 205
ura3, 198, 199, 205, 264
URA3, 198, 199, 207, 213, 222
ura3 lys2, 207
URA5, 199
URF, 71, 72, 74, 86, 87
Uromyces appendiculatus, 418, 419
Ustilago, 534
Ustilago cynodontis, 69
Ustilago maydis, 370, 371, 542
Ustilago zeae, 365
U1 snRNAs, 499

V

var1, 67, 70, 72, 75, 76, 77, 79, 81, 89
Vectors, 135, 208, 463
  integrative, 135, 208, 211
  manipulation, 137
  recovery, 137
Vitamin A, 379
Vitamin B6, 379
Vitamin C, 379
Vitamin D, 380
Vitamins, 378, 379

W

ω-Amino acids, 325
Wangiella, 534
Wangiella dermatitidis, 474
Whittaker, 515, 517
Whole-genome comparisons, 44
Wide domain regulation, 309, 313, 315, 319, 324, 325

X

X-gal, 151, 220, 274
Xanthylic acid, 376

Y

ya, 347
Yarrowia lipolytica, 363, 534, *see also Saccharomycopsis; Candida*
YCp19, 168
YCp630, 208
Yeast, 506, 507, 508, 510, *see also Saccharomyces*
  cloning vectors, 209
  cosmid vectors, 502
  DNA, 222
  enolase gene ENO1, 503
  glycan, 374
  hyphal cell walls, 455
  leu2 structural gene, 503
  plasmids, 167, 168
  telomeres, 173
  transformation, 197, 198
  2-μm plasmid, 503
  YARp (yeast acentric replicating plasmids), 173
  YCp (yeast centromeric plasmid), 135, 174
  YEp (yeast episomal plasmid), 135, 169, 176, 177
  YIp (yeast integrating plasmids), 135, 171, 176, 177, 186, 211
  YRp (yeast replicating plasmids), 135, 177
YEp13, 168, 501
YEp24, pFL1, 205
YEp620, 208, 209, 222
YIp333, 208
YIp5, 168, 208, 436
YIp600, 209
YLP, 177
YLp21, 168
YRMp, 177
YRp610, 208, 209, 211
YRp7, 168, 171
YRp7, 476, 485

Z

Zoosporic fungi, 48
Zygomycotina (zygomycetes), 36, 39, 53, 532, 534
Zygosaccharomyces, 52
Zymolyase, 37, 163
Zymosan, 374

Volume 8 Number 1 January 1992

yeast

ISSN 0749 503X
YESTE3 8(1) 1–78 (1992)

WILEY
Publishers Since 1807
CHICHESTER · NEW YORK · BRISBANE · TORONTO · SINGAPORE

The *YJU3* ORF encodes a polypeptide of 313 amino acids with a predicted molecular weight of 35,549 Da. Computer-aided comparisons of the predicted amino-acid sequence of *YJU3* to sequence databases revealed no sequences with statistically significant similarity to *YJU3*. A null allele of the *YJU3* gene was made by inserting the *LEU2* marker gene within the coding region, at the *Bgl*II site (Figure 2B). The disrupted *YJU3* gene was then used to transform a haploid Leu⁻ strain. Southern analysis of Leu⁺ transformants confirmed the predicted structure of the disrupted *YJU3* gene by showing the presence of restriction endonuclease sites diagnostic of the transplacement. These Leu⁺ transformants grew normally and lacked obvious phenotypic abnormalities, suggesting that *YJU3* is not essential to mitotic growth.

ACKNOWLEDGEMENTS

We gratefully acknowledge Dr John Sgouros (MIPS, Munich, Germany) and Dr Elisabetta Balzi (University of Louvain, Louvain-la-Neuve, Belgium) for their help in sequence analysis. This work was supported in part by grants to A. G. from the Services de la Politique Scientifique: Action Sciences de la Vie and by the Fonds National pour la Recherche Scientifique, Belgium.

REFERENCES

Barnes, W.M., Bevan, M. and Son, P.H. (1983). Kilo-sequencing: creation of an ordered nest of asymmetric deletions across a large target sequence carried on phage M13. *Methods Enzymol.* **101**, 88–122.

Eisenberg, D., Schwarz, E., Komaromy, M. and Wall, R. (1984). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**, 97–120.

Marguet, G., Guo, X.J. and Lauquin, G.J.-M. (1988). Yeast gene *SRP1* (serine-rich protein) intragenic repeat structure and identification of a family of *SRP1*-related DNA sequences. *J. Mol. Biol.* **202**, 455–470.

Robbins, J., Dilworth, S.M., Laskey, R.A. and Dingwall, C. (1991). Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64**, 615–623.

Rothstein, R. J. (1983). One-step gene disruption in yeast. *Methods Enzymol.* **101**, 202–211.

Sanger, F., Nickeln, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.

YEAST VOL. 8: 423–488 (1992)

Foreign Gene Expression in Yeast: a Review

MICHAEL A. ROMANOS, CAROL A. SCORER AND JEFFREY J. CLARE

Department of Cell Biology, Wellcome Research Laboratories, Beckenham, Kent BR3 3BS, U.K.

Received 14 February 1992; accepted 22 February 1992

CONTENTS

Introduction
Transformation and selectable markers
  Auxotrophic selection markers
  Dominant selectable markers
  Autoselection
Episomal vectors
  ARS vectors
  2µ-based vectors
  Regulated copy number vectors
Integrating vectors
  YIp vectors
  Transplacement
  Integration into reiterated DNA
  Transposition vectors
Transcriptional promoters and terminators
  Foreign *versus* yeast promoters
  Glycolytic promoters
  Galactose-regulated promoters
  Phosphate-regulated promoters
  Glucose-repressible promoters
  Other regulated promoter systems
  Selection of novel yeast promoters
  Foreign promoter systems
  Yeast terminators
Factors affecting intracellular expression
  Initiation of transcription
  RNA elongation
  RNA stability
  Initiation of translation
  Translational elongation
  Polypeptide folding
  Post-translational processing
  Stability of intracellular proteins
Secretion of foreign proteins
  Introduction
  Vectors and signal sequences
  Glycosylation
  Protein folding and transport
  Proteolytic processing
  Strategies to improve secretion
Expression in non-*Saccharomyces* yeasts
  Introduction
  Pichia pastoris
  Hansenula polymorpha
  Kluyveromyces lactis
  Yarrowia lipolytica
  Schizosaccharomyces pombe
Physiology of foreign gene expression
  Mechanisms of toxicity
  Generation of low-expressing variants
  Large-scale fermentation and optimization
Concluding remarks
Acknowledgements
References

INTRODUCTION

The yeast *Saccharomyces cerevisiae* has several properties which have established it as an important tool in the expression of foreign proteins for research, industrial or medical use. As a food organism, it is highly acceptable for the production of pharmaceutical proteins. In contrast, *Escherichia coli* has toxic cell wall pyrogens and mammalian cells may contain oncogenic or viral DNA, so that products from these organisms must be tested more extensively. Yeast can be grown rapidly on simple media and to high cell density, and its genetics are more advanced than any other eukaryote, so that it can be manipulated almost as readily as *E.coli*. As a eukaryote, yeast is a suitable host organism for the high-level production of secreted as well as soluble cytosolic proteins.

Most yeast expression vectors have been based on the multi-copy 2µ plasmid and contain sequences for propagation in *E.coli* and in yeast, as well as a yeast promoter and terminator for efficient transcription of the foreign gene (Figure 1). The recen

Correspondence to: M.A. Romanos (tel: 081-658 2211 ext. 641[illegible] fax: 081-650 0725).

0749-503X/92/060423-66 $38.00

rapid expansion in yeast molecular genetics has led to a great increase in our understanding of these components, and as a result there is now a bewildering choice of promoter systems and methods for propagating foreign DNA in yeast. In many cases ingenious new approaches have been employed, for example in increasing the strength of native promoters or the stability of expression vectors. We will attempt to review the choices now available and how they relate to different requirements.

Insertion of a foreign gene into an expression vector does not guarantee a high level of the foreign protein; gene expression is a complex multi-step process and problems can arise at numerous stages, from transcription through to protein stability. In the past, heterologous gene expression has often been treated empirically—a number of host organisms might be tested for successful expression and low yields or failures would be passed over. Frequently, inappropriate conclusions had been made, for example about relative promoter strengths, simply from knowledge of the input vector and the final steady-state level of the foreign protein. However there is now considerable accumulated experience on foreign gene expression in yeast. In many instances this has led to the identification of a particular problem at a specific stage in the chain of events, and then often to its solution. Among eukaryotes yeast offers unparalleled scope for solving such problems on account of the power of classical and molecular genetics combined. We have attempted to draw together examples of this from the literature, and from our own experience. These should be helpful in predicting and solving problems with new genes, and should be applicable, in many cases, to other expression systems, prokaryotic and eukaryotic.

The secretion of foreign proteins which are naturally secreted is often necessary for their correct folding, and is highly advantageous because of the initial purity of the product in the substantially protein-free culture medium. Although there have been several commercial successes using yeast, the area has frequently presented problems, especially for larger proteins. Despite increased understanding of the processes of secretion, the greatest success in improving yields in recent years has been with a classical random mutagenesis approach.

A number of other yeasts have become important host organisms for foreign gene expression because of advantages in promoter strength, secretion efficiency, or ease of growth to high cell density. In the future some of these will often be used in preference to *S.cerevisiae*. We have therefore devoted a substantial section to discussing other yeasts, particularly *Pichia pastoris* in which there are many examples of high-level expression.

An area which is frequently ignored by molecular biologists but which must be considered from the outset in vector design is the physiology of foreign gene expression. This includes the physiology of growth to high cell density and promoter induction, as well as the effect of expressing a foreign protein on host cell metabolism. We have gathered examples of the toxicity of foreign proteins and its effect in causing selection of genetic variants expressing lower yields; in many cases these effects can be controlled. Finally, we will discuss the considerations involved in industrial scale-up and fermenter optimization through examples with different promoter systems of *S.cerevisiae* and with *Kluyveromyces lactis* and *P.pastoris*.

TRANSFORMATION AND SELECTABLE MARKERS

Transformation

The first methods for the transformation of *S. cerevisiae* involved enzymatic removal of the cell wall to produce sphaeroplasts which could take up DNA on treatment with calcium and polyethylene glycol.[24,168] Transformants were then plated out in a selective, isotonic top agar for regeneration of the cell wall. A more convenient method was later developed in which intact yeast cells were made competent by treatment with lithium ions.[189] This method is now widely used despite the fact that it gives lower frequencies; a variation using DMSO increases frequency 25-fold.[167] More recently a third approach, electroporation, has been used, and a highly efficient method has been reported by Meilhoc *et al.*[252]

The process of transformation appears to be somewhat mutagenic, both for the host cell[342] and for the introduced DNA.[72] However the frequency of mutation is low enough that it should not be a major concern here. Also, Danhash *et al.*[83] have reported that transformation induces a heritable slow-growth phenotype in *S.cerevisiae*.

An important factor to consider in foreign gene expression is the frequent wide variation in the productivity of different transformants when 2μ vectors are used. This appears to be due to an unexplained stable variation in plasmid copy number between different transformants.[237,292] Clearly it is therefore important to analyse a number of transformants when optimizing expression.

Table 1. Selectable markers for *S.cerevisisae* transformation.

Marker	Auxotrophic/ dominant	Comments
HIS3	A	
TRP1	A	Selection possible in CAA
LEU2	A	*LEU2*-d for high copy number selection
URA3	A	(a) Selection possible in CAA (b) *URA3*-d for high copy number selection[237] (c) Counter-selection using 5-FOA[37] (d) Autoselection in *fur1* (5-FU-resistant) strain[237,304]
LYS2	A	Counter-selection using α-amino adipate[17,61,121]
S.pombe POT	A	Used in *S.cerevisiae tpi*⁻ host; autoselection in glucose[208]
*Tn903 kan*ʳ	D	Active only in multiple copies unless yeast promoter used; selection using G418[154,200,400]
*Cm*ʳ	D	Only effective using yeast promoter; selection using chloramphenicol in glycerol medium only[153]
*Hyg*ʳ	D	Reference 150
CUP1	D	Level of Cu^{2+}-resistance dependent on gene dosage[159]
HSV *TK*	D	Thymidine/amethopterin/sulphanilamide selection; level of resistance dependent on gene dosage[418]
DHFR	D	Methotrexate/sulphanilamide selection; level of resistance dependent on gene dosage[256,419]

CAA, Casamino acids.

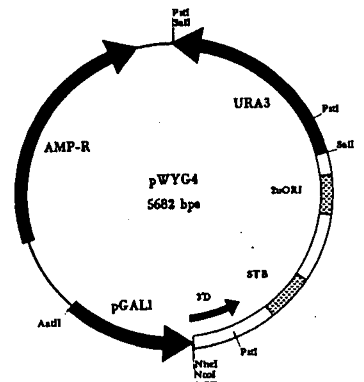
Auxotrophic selection markers

The first and most commonly-used markers for the selection of transformants were *LEU2*, *TRP1*, *URA3*, and *HIS3* used in corresponding mutant strains which are auxotrophic for leucine, tryptophan, uracil and histidine, respectively[24,308,370,382] (Table 1). Such strains are widely available, and some contain non-reverting mutant alleles constructed to give low background rates in transformations. Continued selection requires the use of minimal growth media lacking the relevant nutrient. It is worth noting that *TRP1* and *URA3* vectors can be selected in the presence of acid protein hydrolysates (which lack tryptophan and uracil, e.g. casamino acids) that are often added to semi-defined media in order to enhance growth rates.

A frequently-used variant of *LEU2*, *LEU2*-d,[24] has a truncated promoter and is poorly expressed, so that its selection gives rise to very high plasmid copy numbers.[109] Direct selection of transformants with this marker is inefficient and requires the sphaeroplast method, though low levels can be obtained using the lithium method, if cells are incubated overnight in non-selective medium prior to selection. Vectors which contain both *LEU2*-d and another marker, e.g. *URA3*, can be maintained at either high or low copy number depending on the selection used. Loison *et al.*[237] used an expression vector containing these markers for the schistosomal antigen P28-I and obtained product at 3% of total cell protein (t.c.p.) in uracil-deficient medium or 25% in leucine-deficient medium. Loison *et al.*[237] also constructed a promoter-defective *URA3* allele, *URA3*-d, which gives high copy number in uracil-deficient medium.

URA3 and *LYS2* are particularly versatile in that there are also methods for counter-selection of the marker. Thus *ura3*⁻ cells can be selected for their resistance to the toxic antimetabolite 5-fluoro-orotic acid,[37] and *lys2*⁻ cells can be selected for resistance to α-aminoadipic acid.[17,61,121] These methods can be

a)



b)

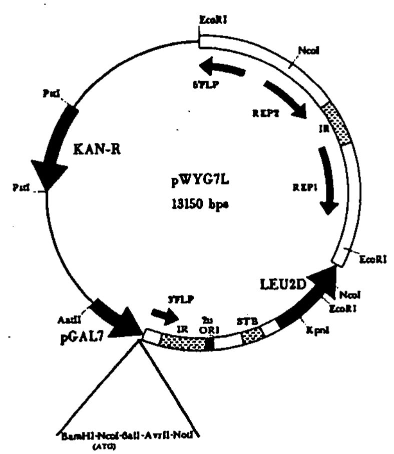


Figure 1. 2μ-based expression vectors for *S.cerevisiae* based on galactose-inducible promoters. (a) pWYG4 has the 2μ *ORI-STB* elements, the *GAL1* promoter, and the 2μ *D* gene terminator; an *Nco*I cloning site containing the ATG is used to insert foreign genes. (b) pWYG7L has intact 2μ *ORI, STB, REP1* and *REP2*, the *GAL7* promoter, and uses the *FLP* terminator; foreign genes are inserted in the polylinker with their 5' ends at the *Bam*HI or *Nco*I site.

used either to select mutations in prototrophic strains or to select for plasmid loss in transformants.

Dominant selectable markers

Dominant markers are useful in that they increase the range of host strains that can be tested to include prototrophic and industrial strains of *S.cerevisiae*, and can be used for selection in rich medium. Since most strains are sensitive to the aminoglycoside antibiotic G418, the G418-resistance marker, encoded by the *E.coli Tn903* transposon, can be used, though it is inefficient in direct selection of transformants.[200,400] Frequencies are acceptable provided transformants are incubated overnight in non-selective medium prior to plating out on G418 agar.[23] Use of glycerol rather than glucose as the carbon source during selection reduces the number of untransformed G418-resistant mutant colonies that arise.[178] Multiple copies of the *Tn903* G418-resistance marker are needed to confer resistance in yeast cells unless a yeast promoter is used, in which case it is active in single copy.[154] It may be prudent to omit selection with antibiotics which affect ribosomes, such as G418, during induction of expression vectors due to the possibility of increased amino acid misincorporation. Other antibiotic-resistance markers that have been used successfully in yeast are hygromycin B[150] and chloramphenicol-resistance.[153]

Copper-resistance in yeast is conferred by multiple copies of the *CUP1* gene and can therefore be used as a dominant marker in sensitive (*CUP1*^{s}) strains.[122] It has proved useful on multi-copy vectors, particularly with industrial strains.[159] Two other markers which can be used for vector copy number amplification by increased drug selection are the herpes simplex virus thymidine kinase gene[418,419] and dihydrofolate reductase.[256]

Autoselection

A number of 'autoselection systems' have been devised to ensure that plasmid selection is maintained, irrespective of culture conditions. Bussey and Meaden[50] showed that expression of a cDNA encoding the yeast killer toxin and immunity gene could be used for self-selection of transformants of laboratory or industrial yeasts since plasmid-free cells are killed by plasmid-containing cells. Another system has used the *Schizosaccharomyces pombe* triosephosphate isomerase gene to stabilize plasmids in *S.cerevisiae* cells lacking the active gene, during growth on glucose.[208]

Loison *et al.*[236] used *ura3 fur1* strains as hosts for plasmids containing the *URA3* gene. These are non-viable since they are blocked both in the *de novo* and salvage pathways for uridine 5'-monophosphate synthesis; maintenance of a *URA3* plasmid is then obligatory for viability even in uracil-containing media. Since the double mutant without plasmid is non-viable, the transformant was obtained by mating a *fur1* strain with a *ura3* strain containing the *URA3* plasmid, and selecting the plasmid-containing *ura3 fur1* progeny. Subsequently, Loison *et al.*[237] were able to directly isolate spontaneous *fur1* mutants from *ura3* cells transformed with a *URA3* plasmid by selecting for resistance to 100 μg/ml 5-fluorouracil and then screening resistant colonies for resistance to 300 μg/ml 5-fluorouridine. We have found that it is possible to generate stable *URA3* transformants by a single direct selection on 1·3 mg/ml (10 mM) 5-fluorouracil to which only *fur1* mutants should be resistant[304] (Figure 2).

EPISOMAL VECTORS

Extra-chromosomal replicons are based either on plasmids containing yeast autonomous replication sequences (*ARS*),[54] which function as origins of replication, or on the native 2μ circle of *Saccharomyces*.[24] A comprehensive listing of both episomal and integrating vectors has been compiled by Parent *et al.*[281]

ARS *vectors*

ARS vectors are present in multiple copies per cell (1 to 20) but are mitotically highly unstable, plasmid-free cells accumulating at a rate of up to 20% per generation without selection, due to inefficient transmission to daughter cells during cell division.[264] Even when grown under selection the proportion of plasmid-containing cells can be very low, giving a correspondingly low average copy number. *ARS* vectors can be stabilized by the addition of yeast centromeric sequences (*CEN*), but the copy number is then reduced to 1 or 2 per cell.[75] In practice *ARS* vectors are hardly ever used for foreign gene expression, and *ARS/CEN* vectors are only used where low-level expression is desired.

2μ-based vectors

By far the most commonly-used expression vectors are *E.coli*-yeast shuttle vectors based on 2μ.[8,44,281] 2μ is a 6·3 kb plasmid present in most

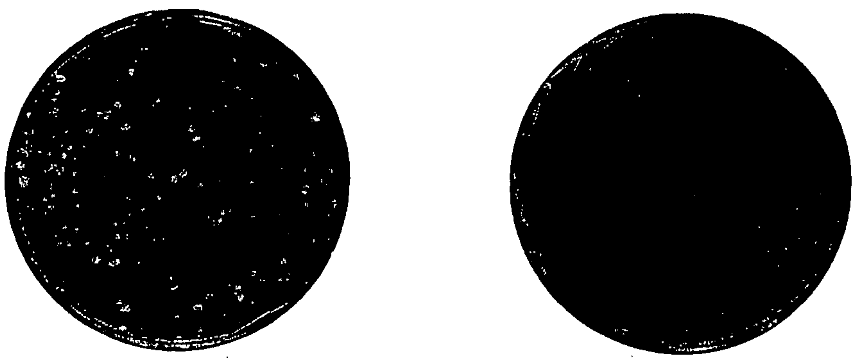


Figure 2. Stability of a *URA3* vector in wild-type and 5-fluorouracil-resistant strains. A β-galactosidase expression vector (pWYG4-*lacZ*) was introduced into strain KY117 and spontaneous mutants resistant to 10 mM-5-fluorouracil were selected.[304] Wild-type (1) and mutant (2) transformants were grown for ten generations in non-selective inducing medium, then plated out on non-selective plates containing XGal and galactose to assay for β-galactosidase expression. White colonies, indicating plasmid-loss, were not present with the mutant strain. This indicates autoselection of the vector in the 5-fluorouracil-resistant strain.

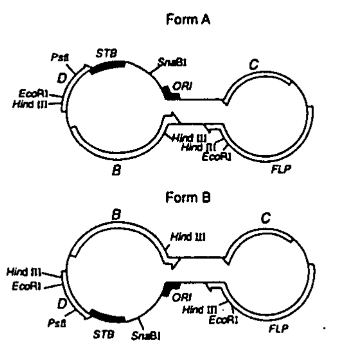


Figure 3. The two forms of the 2μ plasmid. *Cis*-elements are shown as filled boxes and genes by open boxes, inverted repeat regions are aligned; see text for a detailed description.

Saccharomyces strains at about 100 copies per haploid genome[128,265,395] (Figure 3). The plasmid encodes four genes: *FLP* (or *A*), *REP1* (or *B*), *REP2* (or *C*) and *D*. In addition, 2μ contains an origin of replication (*ORI*, which behaves as a typical *ARS* element), the *STB* locus (required *in cis* for stabilization), and two 599 bp inverted repeat sequences. *FLP* encodes a site-specific recombinase which promotes flipping about the FLP recombination targets (FRT) within the inverted repeats, so that cells contain two forms of 2μ, A and B.

Despite the fact that it confers no known phenotype and may indeed be slightly disadvantageous to the host cell,[129,251,397] 2μ is stably inherited; plasmid-free cells arise at the rate of 1 in 10^4 per generation.[129] It achieves this using two mechanisms: (1) by partly overcoming the strong maternal bias in plasmid transmission, and (2) by amplification to correct fluctuations in copy number caused by inefficient transmission.

Efficient segregation depends on having the *STB* locus *in cis* and the *REP1* and *REP2* gene products.[211] Amplification overcomes the host regulation which restricts each replication origin to one initiation per cell cycle; it appears to depend on the inverted repeat sequences and the *FLP* gene product. According to the model proposed by Futcher,[128] FLP promotes recombination between replicated and unreplicated DNA so that inversion occurs and two replication forks can follow each other around the circle. Replication terminates after a second recombination, and further recombination generates multiple 2μ monomers.

The simplest 2μ vectors contain the 2μ *ORI-STB*, a yeast selectable marker, and bacterial plasmid sequences (e.g. pWYG4, Figure 1), and are used in a $2\mu^+$ host strain which supplies REP1 and REP2 proteins.[211] *ORI-STB* expression vectors are the most convenient to use routinely in the laboratory

due to their small size and ease of manipulation. They are tenfold more stable than *ARS* plasmids, being lost in 1 to 3% of cells per generation in non-selective conditions, and are present in 10 to 40 copies per cell. In 2μ-free cells such plasmids behave as *ARS* vectors. Commonly-used *ORI-STB* vectors contain a 2·2 kb *Eco*RI fragment or a 2·1 kb *Hin*dIII fragment from the B form of 2μ, each having one inverted repeat. In order to limit recombination with 2μ, the inverted repeat can be removed, but it is important not to remove adjacent *STB*-distal sequences since these appear to have an important role in protecting *STB* from transcriptional inactivation.[266] It should be noted that shuttle vectors containing inverted repeats will exist as a variety of recombinants with the native 2μ.[139] In addition, there appears to be some competition between exogenous 2μ vectors and native 2μ such that the copy number of both is depressed.[130]

More complex 2μ-based shuttle vectors contain the *REP1* and *REP2* genes in addition to *ORI-STB* and can therefore be used in 2μ-free host strains.[8] They are more cumbersome than *ORI-STB* vectors but stabler and more suitable for scale-up. Many examples of this type of vector are disrupted in *D* (e.g. pJDB248), and some in *FLP* (e.g. pJDB219) so that they cannot flip or amplify in 2μ-free cells. Nevertheless *FLP*⁻ plasmids can reach high copy numbers in 2μ-free cells, presumably through asymmetric segregation. In $2\mu^+$ strains they can amplify independently if they have two inverted repeats, or following integration into 2μ if they have only one repeat. Recently, vectors have been developed which have all functional regions of 2μ intact, by insertion of the foreign DNA at the unique *Sna*BI site between *ORI* and *STB*.[32] These are highly stable, but still 20- to 80-fold less stable than native 2μ, possibly due to low-level transcription through *STB*. In future similar vectors should find wide use in foreign gene expression.

Vectors free of bacterial DNA may be advantageous in foreign gene expression in relation to food and drug regulatory authorities. They can be made in one of two ways: (1) by targetted integration of an expression cassette into native 2μ, or (2) by use of a shuttle vector which can remove bacterial sequences *in vivo* by excisional recombination. Using the latter approach, Chinery and Hinchcliffe[67] constructed 'disintegration vectors' which used the FLP/FRT system to excise bacterial DNA inserted at the *Xba*I site of one of the inverted repeats. Vectors with inserts at the unique *Pst*I site in *D* or the *Sna*BI site between *STB* and *ORI* were extremely stable, though it is not clear whether stability would be retained upon insertion of a high-level expression cassette.

A number of ultra-high-copy number vectors (e.g. pWYG7L, Figure 1) are based on pJDB219 which contains the entire 2μ B form cloned in the bacterial plasmid pMB9 with disruption of *FLP*.[24] The selectable marker is *LEU2*-d whose use results in very high copy numbers (200–400 copies per haploid genome); the fraction of cells with the highest plasmid copy number are constantly selected, resulting in reduced growth rate. Selective growth or use of other markers gives a more normal copy number of about 50 per cell.[109] pJDB219 is best used in 2μ-free cells since it can undergo FLP-mediated recombination with resident 2μ, leading to loss of the foreign DNA but retention of the *LEU2*-d marker.[98] In 2μ-free cells pJDB219 is very stable due to its high copy number, making it suitable for large-scale culture.[130,397] *FLP*⁺ versions of pJDB219, e.g. pXY, have greater stability in non-selective medium.[130] Since the entire sequence of pMB9 has not been determined, it may be more convenient to use a variant of pJDB219 based on pBR322, e.g. pC1/1.[385]

Vectors can have very high copy number even in non-selective conditions if extra FLP recombinase is supplied. For example, induction of FLP from a single integrated copy of the gene under control of the *GAL1* promoter leads to a shift in plasmid copy number from <50 to 200–400.[8] Alternatively, the autoselective high-copy *URA3*-d marker can be used in a *fur1* host strain.[237]

Regulated copy number vectors

Foreign gene expression could be regulated by inducing an increase in vector copy number. Two types of yeast episomal vectors with regulated copy number have been described: (1) vectors with regulatable centromeres, and (2) 2μ vectors in cells with inducible FLP (see above). The first type depends on the observation that *CEN* elements can be inactivated by transcription. Chlebowicz-Sledziewska and Sledziewski[69] constructed vectors containing the glucose-repressible *ADH2* promoter adjacent to *CEN3* and either an *ARS* or 2μ *STB-ORI*. In the *ARS* vectors copy number could be increased from 1–2 to 5–10 by a switch from glucose to ethanol as the carbon source. In the *ORI-STB* vectors copy number could be increased from 1–2 to about 100 and the vector was very stable. Such vectors could be used to increase the degree of regulation in expressing toxic proteins.

EGRATING VECTORS

vectors

ıromosomal integration offers a more stable native to episomal maintenance of foreign A. In *Saccharomyces* integration normally .rs by homologous recombination.[278] Integrat- 'ectors (YIp) contain yeast chromosomal DNA .rget integration, as well as a selectable marker bacterial replicon. Vectors are usually digested unique restriction site in the homologous DNA ıis promotes high efficiency transformation and :ts integration. Such single cross-over inte- ion results in a duplication of the chromosomal :t sequence, so that the vector can subsequently · out' by excisional recombination. Neverthe- the integrants are quite stable, the typical rate ector loss being <1% per generation in the nce of selection.[168]

ɔr convenience vector integration can be tar- :d to the chromosomal mutant allele of the :tion marker used. However, continued selec- of the resulting transformants is ineffective, : the duplicated DNA can be excised and a -type marker retained. In contrast, continued :tion is effective where integration is targetted vhere.

'hen high DNA concentrations of integrating ors are used in transformations, tandem multi- / inserts can result, presumably due to repeated mbination events.[277] Multi-copy integrants are :ively stable and have been used, for example, in : dosage studies.[57]

ısplacement

n alternative type of integration, transplace- .t, makes use of double homologous recombi- on to replace yeast chromosomal DNA, resulting stable structure without duplications.[313] Where ıble single-copy transformant is required this is method of choice. Transplacement vectors con- the exogenous DNA and selection marker ked by yeast DNA homologous to 5' and 3' ons of the chromosomal DNA to be replaced. ır to transformation the vector is digested with riction enzymes which liberate the transplacing ment with 5' and 3' homologous ends. The fre- ncy of transformation is low so that the sphaero- t method is usually used, and the chromosomal cture of the transformants must be checked notypically and by Southern blot analysis.

Integration into reiterated DNA

A number of strategies based on integration into reiterated chromosomal DNA have been used to generate stable multi-copy integrants. At present the best results in terms of copy number and expression appear to be using integration into the ribosomal DNA (rDNA) cluster. The rDNA cluster consists of about 140 tandem repeats of a 9·1 kb unit on chromosome XII. Lopes *et al.*[238,239] have constructed an integrating vector, pMIRY2, containing a portion of the rDNA unit and the *LEU2*-d marker. Transformation with pMIRY2 digested at *Sma*I or *Hpa*I gave Leu^+ transformants with 100–200 copies integrated at a non-transcribed spacer of the rDNA locus. Use of the *LEU2*-d or other promoter-defective markers was important for isolation of high copy integrants. The transformants were highly stable, 80–100% of the integrated copies being retained after 70 generations, and the levels of foreign protein produced using the *PGK* promoter were as high as from 2μ vectors. This approach could also be used as an alternative to episomal vectors in species where none has been found.

Another reiterated DNA that can be used as a target for integration is the transposable element *Ty* which is present in 30–40 copies per genome in most *Saccharomyces* strains. Kingsman *et al.*[216] described the use of a transplacement vector targetted to replace *Ty*, whose copy number could be amplified using the *LEU2*-d selection marker. Levels of interferon produced from such amplified transformants were several times higher than from single-copy *ARS/CEN* vectors but almost ten-fold less than with 2μ vectors. Jacobs *et al.*[193] used similar vectors in order to co-express the different forms of hepatitis B surface antigen (HBsAg). They obtained transformants containing from one to several copies of the vector without the need for amplification. The multi-copy transformants consisted mainly of the transplacing DNA at a single *Ty1* locus and arose by an unexplained mechanism. In order to co-express the different forms of HBsAg in the desired ratio, **a** and α strains were transformed with each vector, multi-copy integrants selected, and diploids made. Stable diploids with a total of 10 copies were made but expression levels appeared to be low.

More recently Shuster *et al.*[345] have used vectors that integrate by single cross-over into δ elements, which exist either alone or as part of *Ty* throughout the *S.cerevisiae* genome. They constructed a vector expressing the *E.coli lacZ* gene with the *LEU2* and *CUP1* markers. Leu^+ transformants were selected

following transformation with vector digested by *Xho*I, which cuts in the δ element within the *LEU2* marker fragment. These were then selected for copper-resistance to isolate multi-copy integrants which yielded up to ten-fold the β-galactosidase level of single-copy strains. Integration into δ, coupled with crossing of haploid integrants, has been used to generate a 20-copy strain for efficient secretion of nerve growth factor.[322]

Transposition vectors

A different approach to multi-copy integration is the use of *Ty* transposition vectors analogous to retroviral vectors for mammalian cells. *Ty* transposes *via* a full-length transcript which is encapsidated into virus-like particles and reverse transcribed to DNA, which can then integrate at multiple sites. In transposition vectors a regulated promoter, e.g. *GAL1*, is used in place of the *Ty* δ promoter to generate a transcript encompassing the foreign gene and selectable marker. Transcription termination signals must be removed from the marker gene for the full-length transcript to be produced. The whole unit is placed on an episomal vector for the initial transformation, but can be lost following induction of transposition. Boeke *et al.*[38] obtained relatively low copy number integrants (1–10) by this method, but it may prove feasible to use the *LEU2*-d gene for amplification. Although this method is not fully developed it could be useful in a variety of yeasts, since *Ty* encodes all the functions required for transposition.

TRANSCRIPTIONAL PROMOTERS AND TERMINATORS

Foreign versus *yeast promoters*

The expression of foreign genes in yeast was examined as soon as transformation procedures became available. The first study was of the rabbit β-globin gene which was found to give rise to aberrant transcripts in which the introns were not spliced.[25] In a few cases the foreign transcript was correctly initiated, e.g. with zein,[228] but in general foreign transcriptional promoters were found to give aberrant initiation, e.g. *Drosophila ADE8*,[161] or were totally inactive, e.g. herpes simplex virus thymidine kinase.[217] Thus for efficient transcription of foreign genes the use of yeast promoters with cDNAs was soon found to be essential. The first published example was the use of a 1500 bp fragment 5' of the *ADH1* gene for efficient intracellular expression of leukocyte α-interferon.[172]

Yeast mRNA promoters consist of at least three elements which regulate the efficiency and accuracy of initiation of transcription:[369] upstream activation sequences (UASs), TATA elements, and initiator elements. Many also contain elements involved in repression of transcription. UASs, which have some similarities to mammalian enhancers, determine the activity and regulation of the promoter through specific binding to transcriptional activators (e.g. *GAL4* and *GCN4*), and act at variable distances 5' to the initiation site. Some UASs have been mapped to short regions of DNA, e.g. the *GAL1-GAL10* UAS to a 108 bp intergenic region containing four short sequences of dyad symmetry. These sequences (17–21 bp) are necessary and sufficient for binding of the *GAL4 trans*-activator and for galactose-regulation, and act synergistically. UASs of some constitutively-expressed genes contain poly(dA-dT) tracts which probably activate transcription by affecting nucleosome structure.[369] TATA elements (consensus TATAA) are found 40 to 120 bp upstream of the initiation site, in contrast to the more rigid distance of 25 to 30 bp in higher eukaryotes, and provide a window within which initiation can occur.[369] The initiator element, which is poorly defined, directs mRNA initiation at closely adjacent sites. Yeast promoters may be highly complex, extending over 500 bp, containing multiple UASs and negative regulatory sites, and multiple TATA elements associated with different initiation sites.[369]

Most promoters are regulated to some extent, but the most powerful, glycolytic promoters are poorly regulated. This makes them undesirable for use in large-scale culture, where there is more opportunity for the selection of non-expressing cells, and unsuitable for expressing toxic proteins. In such cases it is preferable to use a tightly-regulated promoter so that the growth stage can be separated from the expression stage. Despite a severe limitation in efficiency with multiple copies, *GAL1* has been the most commonly-used regulated promoter. However, there is now a large variety of native or engineered promoters (Table 3); the right choice is critical for any one application, especially where a process is to be scaled-up.

Glycolytic promoters

The first promoters used were from genes encoding abundant glycolytic enzymes, e.g. alcohol dehydrogenase I (*ADH1*),[172] phosphoglycerate

Table 2. *S.cerevisiae* vector systems.

Vector	Yeast sequences	Transformation frequency[a]	Copy no. /cell	Mitotic instability[b]	Reference
A. Integrating					
YIp	Homologous DNA	10^2	≥1	0·1%	168
Transplacement	Homologous DNA	10	1	stable	313
rDNA integrating	rDNA	n/a[c]	100–200	stable	239
*Ty*δ	*Ty*δ DNA	n/a	≤20	stable	322, 345
B. Episomal					
Replicating (YRp)	*ARS*	10^4	1–20	20%	264
Centromere (YCp)	*ARS/CEN*	10^4	1–2	1%	75
2μ-based (YEp)	*ORI, STB*, in 2μ$^+$ host (YEp13)	10^4	25	2·8%	130
	ORI, STB, REP1, REP2, FLP in 2μ^0 host (pJDB248)	10^4	50–100	0·6–1·8%	130
	ORI, STB, REP1, REP2, in 2μ^0 host (pJDB219)[d]	n/a	200	0·26%	130
	ORI, STB, REP1, REP2, D, FLP in 2μ^0 host (pJB205)	n/a	50–100	0·20%	32
Regulated copy no.	*ORI, STB, ADH2-CEN3*	10^4	3–5 to 100[e]	Stable[f]	69

[a]Transformants per μg DNA using sphaeroplast method.
[b]Plasmid-free cells arising per generation during non-selective growth.
[c]Data not available.
[d]Increased copy number using *LEU2*-d selection.
[e]On change from glucose to ethanol.
[f]85% after 30h non-selective growth in log-phase.

kinase (*PGK*),[173,383] or glyceraldehyde-3-phosphate dehydrogenase (*GAP*).[177] These were at first thought to be constitutive but were later found to be induced by addition of glucose, e.g. expression of α-interferon using the *PGK* promoter was induced 20- to 30-fold by addition of glucose to a culture grown on acetate as carbon-source.[383] Glycolytic promoters are the most powerful of *S.cerevisiae*, for example *PGK* mRNA accumulates to 5% of total. Despite their poor induction ratio, *ADH1*, *PGK* and *GAP* vectors have been used extensively in the laboratory, and in some cases industrially.[250]

The *PGK* promoter has been studied in some detail: it extends over 500 bp, contains a complex UAS at -473 to-422 relative to the initiation site, a heat shock regulatory element, and other features contributing to accurate and efficient initiation.[215] In contrast, less is known about other glycolytic promoters, such as *GAP*. There are three *GAP* genes, of which *GAP491* or *TDH3* is the most highly-expressed and whose promoter has been used successfully to express a number of proteins.[311] The full-length promoter extends over approx. 700 bp,[33] but there have been a number of reports of smaller fragments having full promoter activity, e.g. a 198 bp fragment.[169] However, it appears that the activity of shorter *GAP491* promoters is dependent on bacterial vector sequences.[311]

Table 3. *S.cerevisiae* promoter systems.

Promoter	Host genotype	Strength	Regulation[a]	Reference
A. Native				
PGK[b], *GAP*, *TPI*	wt	++++	≤20-fold induction by glucose (*PGK*)	311, 383
GAL1	wt	+++[c]	1000-fold induction by galactose	202
ADH2	wt	++[d]	100-fold repression by glucose	291
PHO5	wt	+/++	200-fold repression by P_i	169
PHO5	*pho4*^{ts} *pho80*	+	50-fold induction by shift 37°C to 24°C	223
CUP1	wt	+	20-fold induction by Cu^{2+}	113
*MF*α1	wt	+	Constitutive in α cells	42
*MF*α1	*sir3*^{ts}	+	10^5-fold induction by shift 37°C to 24°C	42
B. Engineered				
PGK/α2 operator	*sir3*^{ts}	++++	100-fold induction by shift 37°C to 24°C	398
TPI/α2 operator	*sir3*^{ts}	++++?	> 50-fold induction by shift 37°C to 24°C	352
GAP/GAL	wt	+++?	150–200-fold induction by galactose	35
PGK/GAL	wt	+++?	Induction by galactose	77
GAP/ADH2	wt	+++[d]?	Repressed by glucose	78
GAP/PHO5	wt	++++?	Two- to five-fold repressed by P_i	169
CYC1/GRE[e]	Expressing glucocorticoid receptor	+++?	50–100-fold by deoxycorticosterone	332
PGK/ARE[f]	Expressing androgen receptor	++++	Several 100-fold by dihydrotestosterone	293

[a]Induction ratios are reporter-dependent.
[b]5% of mRNA in single copy.
[c]1% of mRNA in single copy but severely limited in multiple copies. Improved by *GAL4* over-expression.
[d]Improved by *ADR1* over-expression.
[e]Glucocorticoid responsive element.
[f]Androgen responsive element.

Galactose-regulated promoters

The most powerful tightly-regulated promoters of *S.cerevisiae* are those of the galactose-regulated genes *GAL1*, *GAL7*, and *GAL10*, involved in metabolizing galactose. Galactose-regulation in yeast is now extremely well-studied and has become a key model system for eukaryotic transcriptional regulation (reviewed in reference 201). Many genes are involved in regulation of *GAL* promoters but the central interaction is between the *trans*-activator encoded by *GAL4*, the repressor encoded by *GAL80*, and the *GAL* UAS (Figure 4). Binding of GAL4 protein to the UAS is necessary for induction; GAL80 protein binds GAL4 and acts as a repressor unless galactose is added. The regions of GAL4 protein that bind GAL80 and the UAS have been defined, as have the structural features of the different *GAL* promoters.

GAL1, *GAL7* and *GAL10* mRNAs are rapidly induced > 1000-fold to approx. 1% of total mRNA on addition of galactose.[363] The promoters are

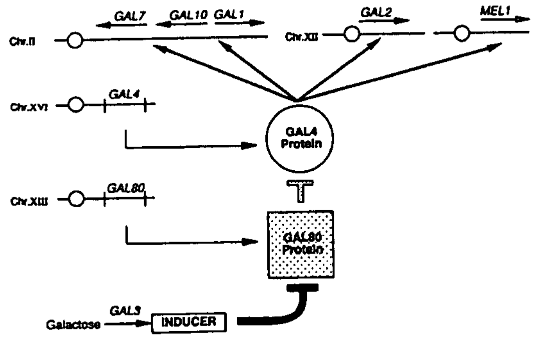


Figure 4. Galactose regulation in yeast. The genes involved in regulation and metabolism and their chromosomal location are shown. Stimulation is indicated by bold lines with arrows and inhibition by lines with bars.

ongly repressed by glucose, so that in glucose- ɔwn cultures maximal induction can only be hieved following depletion of glucose. Galactose- luctions can be carried out in one of three ways: by growing the culture on a non-repressing, non- lucing ('neutral') sugar, when very rapid induc- n follows addition of galactose; (2) by growing culture in glucose medium but removing the ıcose by centrifugation and washing the cells fore resuspension in galactose medium (this leads a lag of 3 to 5 h in induction[1]); and (3) by growing cells in medium containing both glucose and lactose, when the glucose is preferentially metab- ized before galactose-induction can occur. The st two methods are frequently used for small-scale it are impractical in large-scale inductions. It ould be noted that many commonly-used strains ive mutations in the galactose permease gene *AL2*) and are not inducible.

The large background of knowledge of galactose- gulation offers the possibility of manipulating the stem in a number of ways in order to improve its aracteristics for protein production. Three types manipulations that have been carried out are: (1) e over-expression of *trans*-activator; (2) the use of utations in the galactose-regulatory pathway or glucose-repression; and (3) the construction of imaeric galactose-regulated promoters.

GAL4 protein is present in one or two molecules r cell and, moreover, GAL80 repressor is in excess this and is galactose-inducible.[201] Therefore, even ith single-copy promoters, *GAL* transcription is mited by a shortage of GAL4 protein, though other factors also become limiting with multi-copy promoters.[13] With multi-copy expression vectors GAL4 limitation is exacerbated, for example β-galactosidase expression from a *GAL-lacZ* vector increases only two-fold in going from a single-copy to a 2μ vector.[413] Both with single and multi-copy vectors, two- to three-fold increases can be obtained in *gal80* strains, but expression then becomes constitutive.[413] Over-expression of *GAL4*, either by insertion of the gene into the multi-copy expression vector, or by use of a single-copy integrating *ADH1-GAL4* expression vector, increases product levels 2- to 3-fold, but again at the price of losing tight regulation.[13,158] Schultz *et al.*[334] used an integrated *GAL10-GAL4* expression vector in order to over-produce GAL4 protein in a galactose-regulated manner. Transcription from the target expression vector, encoding Epstein-Barr virus (EBV) gp350, was increased 5-fold. We have tested a similar system which used an integrated *ADH2-GAL4* expression vector in order to over-produce GAL4 protein. This system alters the regulation of *GAL* expression vectors so that they are induced by glucose-depletion. Using a constitutive *GAL4* gene, i.e. one that does not interact with GAL80 repressor, it was possible to increase levels of β-galactosidase produced from a multi-copy *GAL1* vector by 2- to 4-fold (M.A.R. and J.J.C., unpublished results).

In order to facilitate galactose-induction of glucose-grown cultures, the use of mutants defective in either global or *GAL*-specific glucose-repression has been examined. In *reg1-1* mutants efficient galactose-induction occurs with galactose to glucose

ratios of 5:1.[248] Hovland *et al.*[181] used a novel selection procedure to isolate glucose-resistant mutants, and found a *reg1* mutant (*reg1-501*) in which efficient induction occurred even with galactose/glucose ratios of 1/100. A *gal1 reg1-501* strain was made which cannot metabolize galactose so that very low levels of galactose (e.g. 0·2%) could be used, in the presence of excess glucose, for efficient induction.[181]

In an attempt to combine the high activity of glycolytic promoters with the tight regulation of *GAL* promoters, hybrid promoters have been constructed where a glycolytic UAS is replaced by a *GAL* UAS. However, the published results do not suggest that these hybrids are more efficient than *GAL* promoters. Bitter and Egan[35] have described the use of a *GAP/GAL* UAS hybrid for the expression of γ-interferon, which is toxic to yeast cells. Insertion of a 55 bp *GAL1-GAL10* UAS fragment between the *GAP* UAS and TATA element conferred galactose-regulation to the promoter. Product levels were significantly higher than with the native *GAP* promoter, due largely to a dramatically increased vector copy number, from 1 to 20–50 per cell. A similar type of hybrid promoter (*PAL*) has been constructed, by replacing the *PGK* UAS with a *GAL* UAS, and used for regulated production of α-interferon and human serum albumin.[77] However, no comparison of the efficiency of *GAL* and *PAL* promoters has been published; it would be expected that all *GAL*-UAS hybrid promoters would be severely limited by GAL4 protein.

Phosphate-regulated promoters

The promoter of the acid phosphatase gene, *PHO5*, is regulated by inorganic phosphate concentration and has been used extensively for foreign gene expression.[169] In one of the earliest studies, a 1·4 kb *PHO5* DNA fragment was used to drive production of α-interferon, which was induced 200-fold by switching to low-phosphate medium.[223] More recently, the structural features and regulation of the *PHO5* promoter have been studied in detail (reviewed in reference 394). The promoter, encompassing about 400 bp of DNA, contains two UASs which are necessary and sufficient for regulation. These contain 19 bp dyad sequences which bind the PHO4 *trans*-activator.

Since the *PHO5* promoter is not very strong, attempts have been made to use the *PHO5* UAS to confer regulation to glycolytic promoters. Hinnen *et al.*[169] constructed a series of *GAP/PHO5* UAS hybrids and tested them in expression vectors for eglin C. Some hybrid promoters yielded up to twice the amount of product as *GAP*, but the induction ratio was 2- to 5-fold compared to 40-fold for *PHO5*. Since no copy number data were available one cannot make conclusions about relative promoter strength.

Kramer *et al.*[223] constructed a host strain temperature-sensitive in the PHO4 *trans*-activator and defective in the PHO80 repressor (*pho4^ts pho80*) to achieve phosphate-independent induction of the *PHO5* promoter regulated by temperature. On lowering the temperature from 35°C to 23°C a 50-fold induction of α-interferon was achieved, but the absolute level was only one tenth that in wild-type induced cells.

Glucose-repressible promoters

Glucose-repression is a global system regulating the expression of a number of genes, including sugar fermentation genes, by the availability of glucose. Yeasts preferentially utilize sugars such as glucose that enter the glycolytic pathway directly. Genes involved in sucrose or galactose metabolism are transcriptionally repressed by glucose. Typical examples of promoters regulated primarily by glucose-repression are those of the *ADH2*, *SUC2* and *CYC1* genes, encoding alcohol dehydrogenase II, invertase and iso-1-cytochrome c, respectively.

The *ADH2* promoter is both powerful and tightly regulated and has been used for foreign gene expression.[291] Since it is repressed over 100-fold by glucose, it can be used for efficient expression of toxic proteins, e.g. insulin-like growth factor I (IGF-I).[343] Deletion analysis has identified a 260 bp region 5' to the initiation site which contains two UASs sufficient for full promoter activity and regulation.[26] UAS1 is a 22 bp inverted repeat which binds the ADR1 *trans*-activator.[105] We have used the 260 bp region, which is readily assembled from synthetic oligonucleotides, in efficient expression vectors (M.A.R. and J.J.C., unpublished results).

In order to maintain repression of *ADH2*, cells must be grown in excess glucose (e.g. 8%) until induction, which is effected by changing to fresh medium containing a non-fermentable carbon source, e.g. ethanol, glycerol, raffinose, etc. Alternatively, *ADH2* can be induced by culturing initially in a lower concentration of glucose (e.g. 1%) which is gradually depleted.[291] Glucose-repressible systems have a potentially serious disadvantage in industrial fermentations: it is difficult to maintain tight glucose-repression under conditions of glucose-limitation, which are required to achieve high cell density.

Irani *et al.*[187] showed that transcriptional factors, including the *ADR1 trans*-activator, became limiting for *ADH2* transcription from multi-copy plasmids. Over-expression of *ADR1* leads to loss of regulation, but *ADH2*-regulated over-expression of *ADR1*, from a single-copy integrating vector, has been used to increase the efficiency of *ADH2* expression systems without loss of regulation.[291] In this way it has been possible to increase expression of β-galactosidase from a single-copy *ADH2-lacZ* vector by 4- to 10-fold. The system usually results in about a 3-fold increase in yield for intracellular proteins and is used for commercial production (J. Shuster, personal communication).

Hybrid glycolytic/*ADH2* promoters have been constructed by transplanting the *ADH2* UAS into *GAP* proximal promoter sequences. Cousens *et al.*[78] fused the *ADH2* UAS containing the 22 bp dyad 320 bp upstream of the *GAP* TATA element, and were able to achieve tightly regulated production of a superoxide dismutase (SOD)-proinsulin fusion to >15% t.c.p. However, no critical comparison of the strength of the *ADH2*, *GAP*, and hybrid promoters has been published.

Other regulated promoter systems

A few other regulated promoter systems are worthy of comment here: these are of interest because their induction is independent of culture nutrients and can therefore be controlled without otherwise perturbing the culture.

A temperature-regulated system based on mating type control has been used by a number of groups. Mutations in the *SIR* genes de-repress the silent mating-type loci, so that a or α-specific genes are not expressed. Repression of the *MFα1* promoter is mediated by the *MATα2* repressor which binds a 31 bp operator sequence. Brake *et al.*[42] first reported the use of a *MATa sir3*ts strain for secretion of hEGF using the α-factor (*MFα1*) promoter. Switching from the non-permissive (35°C) to the permissive (25°C) temperature led to an induction in secreted hEGF from 10 ng/l to 4 mg/l. The α2 operator has been transferred to the strong *ADH2* and *TPI* (triose phosphate isomerase) promoters and shown to confer temperature regulation in a *MATa sir3*ts strain. Sledziewski *et al.*[352] inserted up to four operators between the UAS and TATA elements of the *TPI* promoter and were able to obtain tight regulation (>50-fold induction of β-galactosidase) and full activity. Inexplicably, the hybrid *ADH2* promoters were not tightly temperature-regulated. The system appears amenable to fine-tuning by use of intermediate induction temperatures. A similar system has been developed by insertion of two α2 operators in the *PGK* promoter, achieving a >100-fold induction ratio without any reduction in activity.[398]

A novel type of promoter system uses the ability of mammalian steroid hormone receptors to function as transcriptional activators in yeast. Schena *et al.*[332] constructed a *CYC1* expression vector under the control of three tandem 26 bp glucocorticoid response elements fused upstream of the promoter. In yeast cells also expressing the glucocorticoid receptor, a CAT reporter gene could be induced 50- to 100-fold to high levels by addition of deoxycorticosterone (DOC). The degree of induction was titratable over 1 nM to 10 μM-DOC and was very rapid ($t_{1/2}$ of 7 to 9 mins). A similar system has been developed using a hybrid *PGK* promoter containing androgen response elements.[295] Addition of increasing amounts of dihydrotestosterone induced increasing amounts of β-galactosidase reporter over a several hundred-fold range, and the range could be extended to 1400-fold by varying the copy number of the receptor or reporter genes. In conclusion, steroid-regulated systems appear to be powerful and tightly-controlled, and could be used in yeasts lacking promoters with these qualities (e.g. *K.lactis* and *S.pombe*).

The promoter of the *CUP1* gene, encoding copper metallothionein, has been used in expression vectors.[113] This promoter is tightly-regulated and independent of culture parameters. The concentration of Cu^{2+} ions for induction depends on the copper-resistance of the host strain from 0·01 mM (no *CUP1* gene) to 0·5 mM (>6 copies).

Selection of novel yeast promoters

A number of selection or screening methods have been used in order to identify new promoters from a genomic library.[147,326] Selection procedures can be devised for regulated promoters, e.g. the glucose-repressible promoters of the *GUT2* and *PRB1* genes which encode glycerol-3-phosphate dehydrogenase and vacuolar endoprotease B, respectively.[354,356]

Foreign promoters systems

Foreign promoters not recognized by yeast RNA polymerase can in principle be used provided the cognate RNA polymerase is co-expressed. The bacteriophage T7 RNA polymerase is highly active and has been used in a number of prokaryotic and

eukaryotic organisms. T7 RNA polymerase, ideally with an added nuclear targetting signal, can be expressed from a *GAL1* promoter and can efficiently transcribe DNA in yeast cells.[29,63] However the T7-induced *lacZ* mRNA is not translated in yeast, due either to absence of 5' caps and/or polyadenylation, or to a stable hairpin formed in the 5' region.[29] In animal cells a similar problem has been solved using the encephalomyocarditis virus leader to promote cap-independent translation,[108] but this appears to be inactive in yeast (C.A.S., unpublished results). If the translation problem were solved a yeast T7 system would be very powerful, since *GAL1*-driven expression of the polymerase would act as an amplification step.

Yeast terminators

Yeast transcriptional terminators are usually present in expression vectors for efficient mRNA 3' end formation. Terminators of prokaryotic or higher eukaryotic genes are not normally active in yeast, though there are exceptions such as the *Drosophila ADE8* gene.[161] Efficient termination is probably required for maximal expression: deletion of 'termination' sequences 3' of the *CYC1* gene resulted in longer mRNA and a dramatic reduction in mRNA level.[415]

Transcriptional termination of yeast mRNAs is less well understood than that in bacteria and in higher eukaryotes. Bacterial transcription terminates at the mature 3' end of the mRNA. In higher eukaryotes mRNA 3' end formation involves cleavage and polyadenylation, downstream of the signal AAUAAA, of precursor mRNAs that extend several hundred nucleotides beyond the coding region. Contrary to earlier ideas, it appears that yeast mRNAs follow the same pattern of termination, processing and polyadenylation of pre-mRNA as higher eukaryotes. However in yeast these processes are tightly coupled and occur within a shorter distance, near the 3' end of the gene.[52]

A number of consensus sequences have been implicated as part of the mRNA 'terminator', especially the tripartite sequence TAG..(T-rich)..TA(T)GT..(AT-rich)..TTT[416] and TTTTTATA.[160] The commonly found tripartite motif shows poor conservation and is tolerant of large sequence alterations, suggesting that a general feature such as high AT-content may be critical.[279] However this cannot be sufficient since terminators are frequently absent in AT-rich DNA, and can be unidirectional, implying some sequence specificity. More recent evidence suggests that a variety of different signals, unidirectional and bidirectional, are used in yeast.[188]

Terminators from a number of genes have been used in expression vectors, e.g. *TRP1*,[172] *ADH1*,[385] *GAP*,[310] *MF1*,[42] etc. In order to simplify vector construction, a terminator from 2μ[371] can be used, e.g. the *FLP*[173] (pWYG7L, Figure 1) or *D* gene terminator[304] (pWYG4, Figure 1).

FACTORS AFFECTING INTRACELLULAR EXPRESSION

Initiation of transcription

Gene expression is most frequently regulated at the level of transcription, and it is generally assumed that the steady-state mRNA level is a primary determinant of the final yield of a foreign protein. The mRNA level is determined both by the rate of initiation and the rate of turnover.

In most cases the yield of a foreign protein expressed using a yeast promoter has been much lower than the yield of the homologous protein using the same promoter. Using the *PGK* promoter on a multi-copy vector several proteins accumulate to 1 to 2% t.c.p., whereas with the entire *PGK* gene phosphoglycerate kinase accumulates to over 50%.[62,254] These levels reflect to a large extent the lower amounts of the foreign *versus PGK* transcripts for as many as 20 genes,[62] although there appear to be exceptions.[154] It was suggested that the reduced levels were due to shorter $t_{1/2}$s for the foreign transcripts, but Mellor *et al.*[253] showed that α-interferon mRNA was not unstable but was initiated at a six-fold lower rate. Addition of downstream *PGK* sequences restored the mRNA level to that for *PGK* mRNA, suggesting the presence of a downstream activation sequence (DAS), localized to the first 79 codons, required for maximal transcription. Indirect evidence for a DAS has also been found with the pyruvate kinase (*PYK*) gene.[292] Attachment of the *lacZ* gene to the *PYK* promoter resulted in a 30-fold drop in mRNA molarity, whereas its $t_{1/2}$ decreased only two-fold, consistent with a 15-fold drop in the initiation rate. The putative *PYK* DAS is active in single- and multi-copy vectors, has been localized to nt 500 to 800 relative to the initiating ATG, and appears to be able to functionally replace the *PGK* DAS (A. Brown, personal communication). Evidence for DASs has also been found in the lipoamide dehydrogenase gene,[414] and in *Ty2* DNA.[232]

e 4. Fortuitous transcriptional termination in foreign genes.

:	Size (kb)	AT-content %	RNA analysis[a]	Comments
:tis nid k1 ORF2	3·5	73	Truncated mRNA	Reference 305
nus toxin nent C	1·5	71	Truncated mRNA	≥6 terminators, eliminated by increasing GC-content[b,c,307]
gp120	1·5	60	Truncated mRNA in *P.pastoris* not *S.cerevisiae*	Mutation of T_4ATA_3 gave several longer low-abundance mRNAs Increasing GC-content[b] gave full-length mRNA (R.G.Buckholz, personal communication and C.A.S.)
gp130	1·6	62	Truncated mRNA	(C.A.S., unpublished results)
modium parum p195 : fragment)	0·88	74	Multiple short mRNAs	(C.A.S., M.A.R. and J.J.C., unpublished results)
lciparum '(CDC28 iologue)	0·91	67	No data	No complementation (P.Ross-MacDonald and D.H.Williamson, personal communication)

detectable protein in any example.
:hemical synthesis of the DNA.
milar problem in *P. pastoris* with two bacterial genes (63% and 67% AT) was solved by increasing GC-content (K. Sreekrishna, onal communication).

ince the evidence may now favour the existence)ASs in certain genes it is possible that they will found in many others. There may be some cir- nstantial evidence for this, e.g. multi-copy *GAL7* :s levels of uridyltransferase of >15% t.c.p.,[13] :reas foreign proteins rarely reach this level. It st be emphasized that many other factors could ount for these differences. If DASs are character- l, it may be possible to incorporate them into :tream promoter fragments in order to create re efficient expression vectors. Alternatively, if ,Ss prove to be strongly position-dependent, they ıld be placed within an intron which would be ised prior to translation. If neither of these ions work, then maximal transcription will only be ievable using fusion proteins, for example to PGK. \n alternative to all these approaches is to use a :erent yeast, such as *P.pastoris* which does not)ear to have the same problem.

Yeast expression vectors frequently give rise to merous unexpected transcripts arising from for- tous promoters in bacterial plasmid DNA.[245] It is t known whether they can affect foreign gene expression, but antisense transcripts through the foreign gene can be eliminated by using bidirectional terminators or by judicious juxtaposition with 2μ DNA (e.g. *FLP* terminator).

RNA elongation

The elongation of transcripts is not thought normally to affect the overall rate of transcription, but the yield of full-length transcripts could be affected by fortuitous sequences in foreign genes which cause pausing or termination. These could either act in the same way as natural yeast terminators or else by a different mechanism. We have found evidence of this in a remarkably high proportion of the genes we have examined, especially, but not exclusively, in those with unusually high AT-content. Though not widely recognized, this problem could be a very common reason for low yields or complete failure of foreign gene expression in yeast.

We have listed examples of lack of expression of AT-rich genes in Table 4. The presence of truncated mRNA has not been demonstrated in every case, but the clearest published example is in the expression

of the AT-rich *Clostridium tetani* DNA encoding tetanus toxin fragment C.[307] RNA was analysed from the native gene and from each of three versions of the gene containing progressively more synthetic DNA, starting from the 5' end (Figure 5). The transcript length increased through this series, with only the fully-synthetic gene giving full-length mRNA as the major species. All the short and full-length transcripts were discrete, abundant and polyadenylated. The 1·5 kb *C.tetani* DNA was shown to contain at least six fortuitous polyadenylation sites which were eliminated by increasing GC-content (from 29% to 47%). The partially-synthetic genes gave rise to low levels of run-off translation products, but efficient production of fragment C was only achieved with the fully-synthetic gene.

The problem with AT-rich genes means that analysis of genomes with high AT-contents (e.g. *Plasmodium falciparum* or *Dictyostelium discoideum*) by complementation in *S.cerevisiae* is possibly not worth pursuing as a general strategy. Such studies might be attempted in *S.pombe*, though recent work suggests that the two yeasts have a conserved mechanism for mRNA 3' end formation.[184] More worryingly, there is evidence of the same problem with genes having an average AT-content similar to that of yeast (60%). With the HIV-1 *env* gene, truncated mRNA was found in *P.pastoris* but not *S.cerevisiae*. Mutagenesis of a sequence resembling a polyadenylation consensus inactivated a major polyadenylation site but revealed several weak sites which were removed by increasing GC-content using chemical synthesis (C.A.S. and R.G. Buckholz, in preparation).

The frequency of the problem means that it would be desirable to be able to predict its occurrence, and to solve it without resorting to gene synthesis. Unfortunately it appears to be very difficult to identify polyadenylation sites by searching for consensus sequences, especially in AT-rich DNA. In the case of the fragment C gene, one of the short mRNAs could have been due to a single TTTTTATA sequence, but a number of tripartite motifs were found in regions with no polyadenylation sites.[307]

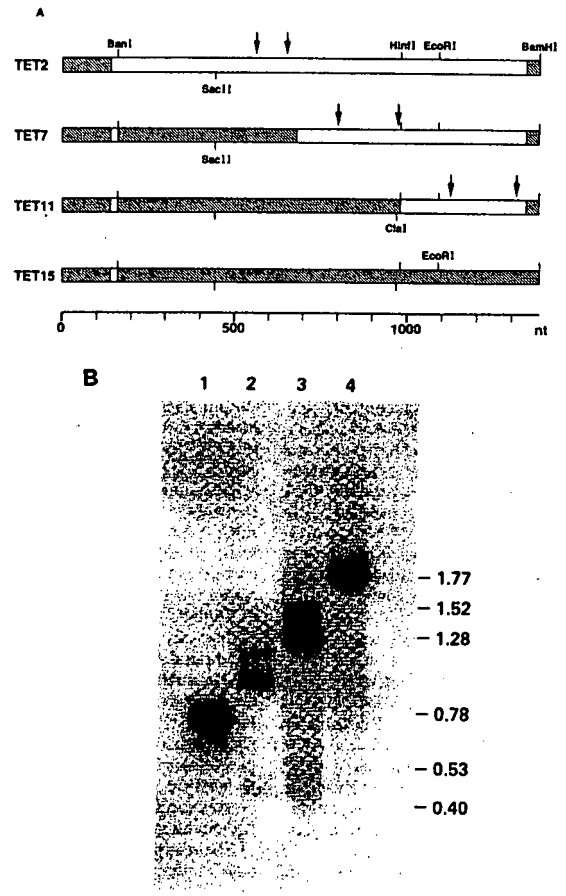
At present the only solution is to increase the GC-content of offending sections of genes by chemical synthesis. The success of more general solutions would depend on whether the problem is due to true premature termination or to uncoupled post-transcriptional processing. In the former case a foreign RNA polymerase which uses different termination signals, e.g. T7 RNA polymerase, might produce full-length mRNA. Unfortunately, the T7 system is not fully developed in yeast (see Transcriptional promoters and terminators). If the problem is one of post-transcriptional processing then it may be necessary to use an extranuclear expression vector, e.g. one based on the linear AT-rich plasmids of *K.lactis* which can also be propagated in *S. cerevisiae*, to segregate the transcripts from nuclear processing enzymes.[365]

RNA stability

The half-lives of yeast mRNAs range from 1 to 100 min and can therefore have a profound effect on the steady-state level (reviewed in reference 47). A careful study of 15 randomly-chosen mRNAs showed two populations, one with longer $t_{1/2}$s (40 to 100 min) and one with shorter $t_{1/2}$s (6 to 20 min).[47] Within each class there was an inverse relation between mRNA length and stability. However, there is some disagreement over this division into two classes.[162] The difference between stable and unstable mRNAs may be that the latter have destabilizing elements, since sequences from the short-lived *URA3* mRNA destabilize *PYK* mRNA.[48] A number of the unstable mRNAs have recently been found to encode ribosomal proteins (A. Brown, personal communication); the foreign *lacZ* mRNA has a $t_{1/2}$ of 27 min.[292]

Insertion of premature stop codons in *URA1*, *URA3* and *PGK* mRNAs has been shown to cause their destabilization, suggesting that ribosome attachment may contribute to mRNA stability.[175,241,283] However the same result was not observed for a *PYK* and *PYK-lacZ* mRNA,[292] suggesting a complex relationship between mRNA stability and translation. Santiago *et al.*[327] found no obvious correlation between ribosome loading and stability for ten different mRNAs.

There appear to be few examples of low mRNA stability being a primary factor in poor yields of foreign proteins. Due to the lack of knowledge of yeast mRNA degradation mechanisms it is impossible to predict whether a foreign mRNA will be stable, though we might expect long mRNAs to be somewhat less stable than short ones. AU-rich sequences in the 3'-ends of some mammalian mRNAs confer instability,[339] but it is not known whether they function in yeast. Where mRNA instability is diagnosed as a problem, overall yield might be improved by (i) using a more powerful promoter, (ii) using a promoter with more rapid induction kinetics, or (iii) chemically synthesizing the gene with altered codons or deleting the 3' untranslated



1 pWYG7–TET2
2 pWYG5–TET7
3 pWYG5–TET11
4 pWYG5–TET15

Figure 5. Fortuitous transcriptional termination in AT-rich tetanus toxin fragment C DNA. (A) Four genes encoding fragment C are shown, with synthetic DNA of increased GC-content (hatched). Approximate 3' ends of transcripts generated in yeast are indicated. (B) Northern blot showing fragment C-specific RNA from these genes: only TET15 gave abundant full-length mRNA.[307]

region[90a] in the hope that instability determinants will be removed.

Initiation of translation

Translational efficiency is thought to be controlled primarily by the rate of initiation. This is affected by the structure of the 5' untranslated leader of a mRNA, and there are examples of poor expression due to the presence of sub-optimal foreign 5' leaders. Using a multi-copy *GAP491* promoter vector to express hepatitis B core antigen (HBcAg) Kniskern *et al.*[218] found that the product accumulated to 0·05% of soluble cell protein when viral 5' and 3' sequences of about 100 nt were retained. Deletion of the 5' viral sequence raised the yield to 26%, while deletion of 5' and 3' viral sequences raised it to 41%, both without altering mRNA levels.

Initiation in eukaryotes is thought to follow a scanning mechanism whereby the 40S ribosomal subunit plus co-factors (eIF2, eIF3, eIF4C, Met-tRNA and GTP) bind the 5' cap of the mRNA then migrate down the untranslated leader scanning for the first AUG codon.[222] Any part of this process, which is affected by the structure of the leader and the AUG context, could limit the initiation rate. Yeast mRNA leaders have an average length of about 50 nt, are A-rich, have little secondary structure, and almost always use the first AUG for initiation (reviewed in reference 70). The length of the untranslated leader does not seem to be critical: although there is a 50% reduction in efficiency when the *PGK* leader is shortened from 45 to 21 nt, further shortening to 7 nt has no effect.[389] Insertion of long runs of Gs is very deleterious, however, causing a complete inhibition of translation; runs of Us have a partial effect possibly caused by interaction with the polyA tail.[390] The most significant single factor is secondary structure: hairpins of $\Delta G = -20$ kcal/mol can drastically inhibit translation of *HIS4* or *CYC1* mRNA in yeast cells,[11,71] in contrast to the much lower sensitivity of translation to leader secondary structure in mammalian cells.

The consensus sequence around the initiating AUG, A/YAA/UAA<u>AUG</u>UCU,[70] in yeast is different from that in higher eukaryotes (CACC<u>AUG</u>G). It has been speculated that the mammalian context reflects an interaction with the sequence GUGG in 18S rRNA. The prevalence of UCU as the second codon may reflect the preference for stabilizing amino acids according to the N-end rule (see below). Although there is a strong bias for A at-3, especially in highly-expressed genes, alteration of the AUG context of mRNAs in yeast has at most a 2- to 3-fold effect, in contrast to the much greater effect in mammalian cells.[11,71]

As a consequence of these considerations it is preferable to avoid the inclusion of foreign non-coding sequences in yeast expression vectors. However, where this is inconvenient it should be possible to predict whether a foreign 5' leader will be deleterious by examining the sequence for secondary structure or runs of Gs or Us. For example, analysis of the sequence 5' to the HBcAg gene, found by Kniskern *et al.*[218] to inhibit translation, shows a number of predicted secondary structures of up to $\Delta G = -35$ kcal/mol (M.A.R., unpublished results).

Translational elongation

Translational elongation is not thought to affect the yield or quality of polypeptide normally, but there is now some evidence that it can become limiting with very high mRNA levels. Codon usage is known to affect the elongation rate and therefore has to be considered as a *potential* factor affecting product yield. In the past it has been discounted on the basis that some genes containing many rare codons, e.g. *lacZ* or HBcAg, are expressed at high levels; however, these gene products are unusually stable and their yield does not reflect overall translational efficiency.

Despite the degeneracy of the genetic code, a non-random codon usage is found in most organisms: whilst the codon usage of most genes reflects the nucleotide composition of the genome, highly-expressed genes show a strong bias towards a subset of codons.[28,338] This 'major codon bias', which can vary greatly between organisms, is thought to be a growth optimization strategy such that only a subset of tRNAs and aminoacyl-tRNA synthetases is needed at high concentration for efficient translation of highly-expressed genes at fast growth rates.[227] Rare codons, for which the cognate tRNA is less abundant, are translated at a slower rate in *E.coli*,[282] but this will not normally affect the level of product from an mRNA since *initiation* is usually rate-limiting. A ribosome finishing translation of one mRNA molecule is most likely to initiate translation of a different mRNA species, unless the original species comprises a large proportion of total mRNA. Thus, the overall rate of translation of an mRNA is not usually affected by a slower elongation rate unless ribosomes become limiting, which would affect all transcripts in the cell.

In contrast to the normal situation, there is evi-ence that codon usage may affect both the yield and uality of a protein when a gene is transcribed to very igh levels. With very high levels of mRNA contain-ıg rare codons, aminoacyl-tRNAs may become miting, increasing the probability of amino acid ıisincorporations, and possibly causing ribosomes ɔ drop off. Indeed, a high misincorporation fre-uency has recently been observed in a foreign pro-ːin produced at very high levels in *E.coli*.[336] The ·resence of many very rare codons has been shown ɔ limit the production of tetanus toxin fragment C ı *E.coli*.[242] In yeast, Hoekema *et al.*[175] showed that ubstitution of a large proportion of preferred odons with rare codons in the 5' portion of the *PGK* ene caused a decrease in expression level. However, hese substitutions were made in the region contain-ıg a putative DAS,[253] possibly contributing to the ffect on expression levels. Translation of *lacZ* ıRNA, which contains a high proportion of rare odons, was shown to be limiting at high levels.[292] 'here was a decrease in ribosome loading, possibly ue to pausing and drop-off, on *lacZ* mRNA and on *'RP2* mRNA, which also has poor codon bias. ːecently, expression of an immunoglobulin kappa hain in yeast was shown to increase 50-fold, with-ut a change in mRNA levels, using a synthetic, odon-optimized gene.[220]

Thus the codon content of a foreign gene may ıfluence the yield of protein where the mRNA is ·roduced at very high levels. This may be more likely ɔ occur on growth in minimal medium, when the cell ·roduces a wide variety of biosynthetic enzymes, ncoded by genes containing rare codons.[338] The ffect on product quality has been difficult to measure ·ut requires further attention since it has important nplications for therapeutic proteins.[336] Proteins ontaining amino acid misincorporations are diffi-ult to separate and may affect the activity and ntigenicity of the product. Since small genes are ow frequently synthesized chemically they may ·e easily and perhaps profitably engineered to con-ıin optimal codons for high-level expression in east.

mRNA secondary structure, in addition to codon sage, may affect translational elongation. Baim *t al.*[10] showed that a mutation which introduced a .airpin loop ($\Delta G = -19{\cdot}6$ kcal/mol) near the 5' end ·f the coding region of *CYC1* mRNA reduced the mount of protein product five-fold. mRNA levels /ere unchanged and analysis of the distribution of *:YC1* mRNA on polysome gradients indicated that ranslation was affected.

Polypeptide folding

During or following translation, polypeptides must fold so as to adopt their functionally-active conformation. Since many denatured proteins can be refolded *in vitro*, it appears that the information for correct folding is contained in the primary poly-peptide structure.[140] However, folding comprises rate-limiting steps during which some molecules may aggregate, particularly at high rates of synthesis and at higher temperatures.[210] There is evidence that certain heat-shock proteins act as molecular chaperones in preventing the formation and accumulation of unfolded aggregates, while accelerating the folding reactions.[140]

Due to the intrinsic nature of polypeptide folding and the low specificity of chaperones, it is very unlikely that foreign cytosolic proteins will accumulate in yeast in non-native conformations, and indeed this is generally the case. When fragments of proteins or fusion proteins are expressed, however, normal folding domains may be perturbed resulting in an insoluble product. Nevertheless, fusion proteins that are insoluble in *E.coli* may frequently be soluble when expressed in yeast, e.g. fusions to β-galactosidase (J.J.C., unpublished results), glutathione-S-transferase (GST; M.A.R., unpublished results), and HBcAg.[23] Insoluble proteins can often be renatured *in vitro* though the techniques for this can be complex and unpredictable.[246]

In contrast to intracellular proteins, naturally-secreted proteins encounter an abnormal environment in the cytoplasm: disulphide bond formation is not favoured and glycosylation cannot occur. Though many secreted proteins are insoluble when expressed intracellularly, e.g. prochymosin,[146] human serum albumin,[294] HIV gp120,[19] some are soluble and biologically-active, e.g. α-interferon,[172] α_1-antitrypsin,[310] tumour necrosis factor,[359] Factor XIIIa.[300] Factor XIIIa is interesting in that it has nine Cys residues, none of which forms disulphide bonds.

In *E.coli*, foreign proteins are frequently insoluble but low temperature has been found to increase solubility in some cases.[331] This may be due to a decreased translation rate or to the fact that hydrophobic interactions, such as occur in aggregates, become less favourable. A dramatic increase in the yield of active, soluble protein has been reported on reducing the rate of induction in *E.coli*.[219] Low temperature or reduced induction rates may increase product solubility in yeast. For example, in the intracellular expression of the bacterial membrane protein pertactin, the proportion of the product that was soluble decreased from 100% to 10% as the expression level was increased from 0·1% to 10% t.c.p.[306]

Table 5. Effect of penultimate amino acid in determining mature N-terminus of intracellular proteins.

	Met removed	Met retained
No N^{α}-acetylation	Pro, Val, Cys	Ile, Leu, Met, Phe, Tyr, Trp, Lys, His, Arg, Gln
N^{α}-acetylation	Gly, Ala, Ser, Thr	Asp, Glu, Asn

Similar considerations apply to the assembly of foreign multimeric proteins. Examples of homo-polymeric assembly include the HBcAg[218] and HBsAg.[387] The latter is inserted into intracellular yeast membranes to form immunogenic 22 nm particles resembling those found in the serum of chronic HBV carriers. In early studies only a proportion of the HBsAg was found to be immunologically-active, and later work showed that mature particles only formed during extraction, when SH residues oxidized to form inter-chain S_2 bonds.[399] More recent work in *P.pastoris* showed that a much higher proportion of HBsAg was correctly folded when produced in a strain growing slowly during induction, suggesting rate-limiting folding steps related to growth rate.[81]

There are few published examples of heteromultimeric assembly in yeast, but a notable one is the co-expression of α and β-globin cDNAs to produce haemoglobin.[396] The yield of fully-assembled, soluble haemoglobin, which incorporated yeast haem, was 3 to 5% t.c.p. Another interesting example is in the simultaneous intracellular expression of heavy and light chains of an IgG directed against ADH I: these were able to assemble in the cytoplasm and partially block ADH I activity *in vivo*, despite the fact that the inter-chain S_2 bonds would not form.[56] Similarly, intracellular expression of a catalytic antibody Fab fragment yielded functional product at 0·1% t.c.p.[39]

Post-translational processing

Amino-terminal modifications of polypeptides are the commonest processing events and occur on most cytosolic proteins (reviewed in reference 209). Two types of events normally occur: removal of the N-terminal Met residue, catalysed by Met amino-peptidase (MAP), and acetylation of the N-terminal residue, catalysed by N^{α}-acetyltransferase (NAT). Both enzymes are associated with ribosomes and act on nascent polypeptides.

The specificities of yeast MAP and NAT (Table 5) have been determined by N-terminal amino acid analysis of mutant iso-1-cytochrome *c* or thaumatin polypeptides,[183,260] and more recently by studies of purified MAP.[60] The specificity of MAP is determined by the size of the penultimate residue. When this is has a radius of gyration of < 1·29 A (Gly, Ala, Ser, Cys, Thr, Pro and Val) the Met is removed, though with the larger residues Thr and Val there is only partial removal if they are followed by Pro. With other penultimate residues Met is retained. These rules appear to be reliable and highly conserved among eukaryotes, so mammalian proteins produced in yeast should have the normal N-terminal residue.

Recently it has been shown that fusions of proteins to the C-terminus of ubiquitin are rapidly processed *in vivo* when expressed in yeast, liberating the mature protein.[321,393] This approach is useful since it can be used to generate a protein with any desired N-terminus (apart from Pro), and appears to significantly increase the product yield in some cases.[101]

The factors governing N^{α}-acetylation are less clear. It appears that Met-Glu/Asp is sufficient for acetylation in eukaryotes.[260] N-terminal Gly, Ala, Ser, Thr and Met-Asn may also be acetylated, though effects of second and third residues make this less predictable.[209] However, there is evidence that the process is conserved among eukaryotes, so that proteins acetylated in mammalian cells might be expected to be acetylated in yeast cells.[156]

In most cases the structure of the N-terminus should not affect biological activity of a protein, but there may be exceptions. For example the response of haemoglobin to physiological modifiers involves the N-terminus, and correct processing of α and β-globins in yeast is therefore advantageous over expression in *E.coli*.[396] N^{α}-acetylation of melano-cyte-stimulating hormone and of β-endorphin is required for full biological activity (reviewed in reference 209).

A variety of other post-translational modifications which are often critical for biological activity appear to be conserved between yeast and higher eukaryotes. The phosphoproteins fos and c-myc are correctly phosphorylated in yeast.[258,325] Myristylation is a co-translational modification of N-terminal Gly important for the membrane targetting of certain proteins, e.g. G proteins, src tyrosine kinases, and retroviral gag proteins, and this also occurs in yeast.[195] Isoprenylation affects an important class of membrane proteins including G proteins and ras proteins:[330] the Cys of C-terminal Cys-X-X-X-CO_2H is isoprenylated, following which the three C-terminal residues are removed and the Cys-CO_2H is methyl-esterified. Mammalian proteins such as H-ras p21 are isoprenylated in yeast.

Stability of intracellular proteins

So far, processes affecting the rate of synthesis of proteins have been considered, but the ultimate yield is equally affected by the rate of degradation. In fact the few examples of very high level expression (>25% t.c.p.) in *S.cerevisiae* are of unusually stable proteins, e.g. SOD,[156] HBcAg,[218] and schistosomal GST.[237] This reflects a difficulty in achieving a very high rate of synthesis of foreign proteins in *S.cerevisiae*, probably at the level of transcription.

Very low yields are obtained with proteins which are naturally short-lived, such as myc,[258] or with some polypeptides which are naturally secreted, such as insulin[366] and hEGF.[385] In some cases fusion to a stable protein has given high-level accumulation in yeast, for example with a hybrid TAT-Ty particle,[40] a SOD-proinsulin fusion protein,[78] fusions to GST (M.A.R., unpublished results), and a variety of peptides fused to HBcAg.[23] Alternatively, secretion has been used to segregate the product from intracellular proteases. In the case of HBsAg particles containing the pre-S2 peptide, proteolytic cleavage occurred at a specific site and could be reduced by using a protease-deficient (*pep4*) strain or by mutation to remove the susceptible region.[190] Where these approaches fail, yields might logically be improved by the following measures: (i) using a more rapidly-induced promoter, (ii) using additional protease inhibitors to minimize degradation during extraction, (iii) inducing at lower temperature, (iv) harvesting cells in the exponential growth phase.

A number of different pathways of protein degradation exist and therefore there are multiple molecular determinants that confer instability (reviewed in reference 96). Unfortunately, we do not know all the determinants nor the relative importance of different pathways. Vacuolar degradation is responsible for non-selective bulk turnover of long-lived proteins (average $t_{1/2}$ approx. 160 h), whereas short-lived proteins ($t_{1/2}$ <2·5 h) are degraded in the cytosol by an ATP-dependent pathway involving ubiquitin.

A number of proteases, activated by the *PEP4* and *PRB1* gene products, are responsible for yeast vacuolar degradation.[170] Mutations in both genes dramatically reduce cellular proteolysis and should also reduce the risk of proteolysis during extraction. *pep4* mutants are widely used, but do not normally appear to offer an advantage in product yield. Indeed a general reduction in protein turnover rate would not be expected to increase the *relative* accumulation of a foreign protein, though it might increase the total protein yield. Vacuolar proteolysis is affected by culture conditions and increases several-fold during N- or C-starvation or in stationary phase.[170]

In the ubiquitin pathway proteins are marked by covalent attachment of ubiquitin, a 76-residue polypeptide, and become substrates for rapid degradation by a cytosolic ATP-dependent proteasome.[96] In addition to short-lived proteins, damaged or denatured proteins conjugate ubiquitin more efficiently and are targetted for degradation.

Varshavsky and co-workers have identified one component of ubiquitin recognition as the N-terminal amino acid (the N-end rule pathway; reviewed in reference 393). The N-end rule was uncovered by the production of β-galactosidase variants containing different N-terminal residues, which were generated by the expression and spontaneous processing of ubiquitin-β-galactosidase fusions; the $t_{1/2}$s of these variants ranged from 2 min to > 20 h. The N-end rule degradation signal actually comprises two determinants: a destabilizing N-terminal amino acid and a proximal internal Lys. Destabilizing residues are classified as Type I (positively charged: Arg, Lys, His), Type II (bulky hydrophobic: Phe, Tyr, Trp, Leu, and Ile in yeast), and Type III (small uncharged: Ala, Ser, Thr), the latter are not destabilizing in yeast. Additionally, the residues Asp and Glu are secondary destabilizing residues since they are substrates for post-translational addition of Arg; Asn and Gln are tertiary destabilizing residues since they can be deamidated *in vivo*. Recognition of a destabilizing N-terminus results in multiple ubiquitination of the internal Lys residue which is the prelude to degradation.

In agreement with the N-end rule, almost all cytosolic proteins with known N-termini have stabilizing amino acids. This might be expected from the striking inverse correlation between the N-end rule and N-termini that are generated by MAP. (In contrast most secreted proteins have destabilizing N-termini.) However, recent data suggest that the N-end rule is only one component, possibly a minor one, of the ubiquitin pathway. Mutations in the yeast gene (*UBR1*) for the Type I and II recognition component dramatically stabilized Arg-galactosidase but did not affect either bulk proteolysis or ubiquitin-dependent degradation.[20] A similar result was obtained using inhibitors of Type I (Arg-Ala) or Type II (Leu-OMe) recognition *in vivo*, resulting in a ten-fold stabilization of Arg- or Leu-galactosidase, respectively, and > 50-fold increases in accumulated protein.[12] Therefore other, undefined, signals are more important in the ubiquitin pathway.

Another proposed signal for rapid degradation is a variable sequence rich in Pro, Glu, Ser, and Thr (PEST) found in the majority of short-lived proteins.[301] Recently, addition of PEST sequences has been shown to destabilize DHFR.[235] It is not clear whether degradation of proteins containing PEST sequences involves the ubiquitin pathway.

SECRETION OF FOREIGN PROTEINS

Introduction

A wide variety of heterologous proteins have been secreted from yeast and this approach offers certain advantages over intracellular production. Although *S.cerevisiae* secretes only 0·5% of its own proteins, this level can be increased several-fold so that a secreted foreign protein can be almost pure in the medium. Many pharmacologically-important proteins are naturally secreted and can often only adopt their correct conformation by folding within the secretory pathway. As a consequence, intracellular expression is often unsuitable since the product may be insoluble and may also have an incorrect N-terminus, although this can be overcome by fusion to ubiquitin (see 'Factors affecting intracellular expression'). Thus secretion is used mainly for the production of correctly-folded, naturally-secreted proteins, but there are, in addition, other instances when secretion may be preferable. For example, some proteins are unstable or toxic when cytoplasmically-expressed and these problems may be circumvented by secretion.

As in higher eukaryotes, protein secretion in yeast is directed by an amino-terminal signal sequence which mediates co-translational translocation into the endoplasmic reticulum (ER). The signal peptide is removed by a signal peptidase. In the lumen of the ER

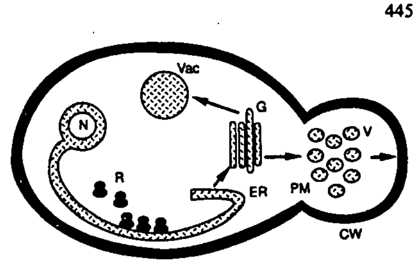


Figure 6. The yeast secretory pathway. Arrows indicate the route taken by proteins through the secretory pathway to either the vacuole or the plasma membrane. N, nucleus; R, ribosomes; ER, endoplasmic reticulum; G, Golgi; Vac, vacuole; V, secretory vesicles; PM, plasma membrane; CW, cell wall.

asparagine-linked glycosyl structures may be added. The signal for the addition of these N-linked sugars is the same for yeast and mammalian glycoproteins (Asn-X-Ser/Thr). O-linked oligosaccharides may also be added. Proteins are then transported in vesicles to the Golgi where modifications to these glycosyl structures take place. These modifications differ from those made by higher eukaryotic cells and, as a result, glycosylation is increasingly regarded as a major drawback to the secretion of therapeutic glycoproteins from yeast. From the Golgi, proteins are packaged into secretory vesicles and are delivered to the cell surface (Figure 6).

Once in the ER, it is probable that a default pathway directs a protein to the plasma membrane unless it contains specific signals to cause retention in the ER or Golgi, or to target it to the vacuole. Therefore it might be thought that if a foreign protein could be targetted to the lumen of the ER, it should be successfully secreted. However, there are a number of stages in the secretory process at which problems may occur. The yeast proteins which assist in folding and disulphide bond formation differ from their counterparts in higher eukaryotes and this may affect folding of foreign proteins. Malfolding can result in retention in the ER and degradation. There are also reports of proteins being retained in the Golgi, again possibly due to malfolding. Retention in the cell wall has also been a problem, especially with larger proteins, although factors other than molecular mass are known to be important. In addition to problems of transport, other undesirable events such as aberrant processing or hyperglycosylation may take place during the secretory process. It should be noted, however, that proteins which are poorly secreted from *S.cerevisiae* may be efficiently secreted from other yeasts, for

ample prochymosin secretion is much more icient from *K.lactis* and *Yarrowia lipolytica* an from *S.cerevisiae* (see 'Expression in non-*ccharomyces* yeasts').

The secretion of heterologous proteins from yeast s been most successful with peptides, many of iich are commercially important, e.g. epidermal owth factor (EGF) and insulin. In high-density ltures hundreds of milligrams of these proteins ay be secreted per litre. The secretion of large pro-ns has proved less predictable, nevertheless there ve been some notable successes.[186,284,388] The ccessful secretion of proteins which are not natur-ly secreted, such as HIV-1 protease has also been ported,[288] but the presence of fortuitous ycosylation sites can cause problems.[307]

ctors and signal sequences

omoters and selectable markers Most commonly-ed secretion vectors are based on high-copy, 2μ asmids. Integrated vectors have been reported to ve higher yields of secreted product than episomal ctors; four integrated copies of a prochymosin pression unit resulted in similar overall expression vels but higher secretion yields than were achieved ith a multi-copy vector.[357] The reason for this fference is unclear. However, yields of EGF, which contrast to prochymosin is efficiently secreted, ere not higher from integrated gene copies.[74] Sakai al.[322] compared secretion of human nerve growth ctor from a 20-copy *Ty* δ-integrant with that from 2μ vector. Levels were three- to four-fold higher –4 mg/l) from the δ-integrant, but the use of a nstitutive promoter (*PGK*) may have led to copy umber reduction with the 2μ vector.

A number of the promoters described above (see 'ranscriptional promoters and terminators') have en used in secretion vectors, but frequently a pro-oter and signal sequence from the same gene are osen e.g. *MFα1*, *PHO5* or *SUC2*. Ernst[110] ported that up to a two-fold increase in somato-edin-C secretion could be obtained by using the eak *CYC1* promoter rather than the moderate tin promoter. Product toxicity may be more acute ith powerful promoters and this may work to duce plasmid copy number; the use of weak or oderate constitutive promoters, or regulated pro-oters which are repressed during early growth hases, may minimize this effect.

The choice of selection marker on a secretion vec-or may be particularly important since culture con-itions may dramatically affect the final yield in the edium. Selection of plasmid-containing cells in a defined medium may result in lower cell density and, in some cases, lower levels of secreted product per cell. Wheat α-amylase secreted from yeast has been shown to reach much higher levels in rich medium than in selective, defined medium.[314] At laboratory scale, it may be preferable therefore to use a dominant marker such as that for G418 resistance.

The nature of signal sequences A classical signal sequence comprises a charged N-terminus, a central hydrophobic core and a consensus sequence for cleavage in the ER by signal peptidase. Some secreted proteins, such as the yeast mating pheromone, α-factor, have additional pro sequences which may aid secretion.

As described below, heterologous proteins may be secreted from yeast using either a foreign signal, often derived from the protein being secreted, or a yeast signal. Since signal sequences are recognized with low specificity in yeast,[205] it could be assumed that foreign signals would work as efficiently as those from yeast, but this is often not the case.

A number of studies aimed at identifying the features of yeast signal peptides have been performed. Ngsee *et al.*[267] analysed mutated signal sequences from the yeast invertase gene, *SUC2*. They concluded that the essential feature of a signal peptide is a hydrophobic core of 6–15 amino acids, which may be interrupted by non-hydrophobic residues. Additionally, many signal peptides were found to contain one or more basic amino acids preceding the hydrophobic core. Small neutral and α-helix-disrupting amino acids were often present in the vicinity of the cleavage site.

Secretion of α-factor was shown to be drastically reduced if the hydrophobic core of the signal was interrupted by a hydrophilic residue, or a proline which disrupts-helical secondary structure.[4] Ngsee and Smith[268] noted that yeast invertase signal sequence is predicted to have an α-helical conformation, whereas that of bovine prolactin has a tendency to form an extended coil. Substitutions in the prolactin signal which increased the probability of it having an α-helical conformation improved its functioning in yeast. Thus there may be more stringent requirements for an α-helical secondary structure in yeast signal sequences compared to those of higher eukaryotes.

Surprisingly, even if the signal peptide is deleted, some proteins, e.g. carboxypeptidase Y[36] and acid phosphatase,[349] may still inefficiently enter the secretory pathway, indicating that the mature protein sequence may also contain features which can be recognized by elements of the secretory pathway.

Use of heterologous signal sequences Early attempts to secrete foreign proteins from *S.cerevisiae* utilized the protein's own signal sequence, e.g. *E.coli* β-lactamase,[303] human α- and γ-interferon (IFN),[173] mouse immunoglobulin heavy and light chains[406] and influenza virus haemagglutinin.[191] However the expression levels were often very low, with only a proportion of the protein being secreted. Processing of the polypeptide was not always correct and, in the case of IFN, degradation of the preprotein took place.[173] Nevertheless, some foreign proteins have been successfully secreted from yeast using their own signal, e.g. *Aspergillus awamori* glucoamylase,[186] barley α-amylases 1 and 2[358] and human serum albumin (HSA).[355] De Baetselier *et al.*[89] have reported the production of up to 3 g/L of active, secreted *Aspergillus niger* glucose oxidase.

The secretion from yeast of *Bacillus amyloliquefaciens* α-amylase[316] using its own signal demonstrated that a prokaryotic signal sequence may also efficiently direct secretion. The signal peptide was correctly cleaved, but the α-amylase was core glycosylated in contrast to the native product. However, the glycosylated enzyme retained its activity.

Foreign signals may also be used to drive secretion of other heterologous proteins—the human gastrin signal sequence has been shown to drive secretion of human α-amylase in *S.cerevisiae*.[328] In general, the heterologous signal sequences which function best in *S.cerevisiae* tend to be those from plant or fungal proteins (e.g. barley α-amylase,[358] *Trichoderma reesei* endoglucanase,[388] *Mucor pusillus* aspartic proteinase[410]).

The leader sequence from the killer toxin *ORF2* gene of *K.lactis*, a yeast closely related to *S.cerevisiae*, has been employed to direct the secretion of human interleukin-1β[15] and diphtheria toxin-hormone fusion proteins.[285] HSA has also been secreted from *S.cerevisiae* using this leader sequence, however it was shown to be less efficient than the native HSA leader or an *S.cerevisiae* signal sequence.[355]

Although there have been some notable successes, the use of foreign leaders often results in intracellular accumulation (e.g. α-IFN,[173] α-1-antitrypsin[53]), and it must be concluded that the yeast secretory pathway has slightly different requirements from higher eukaryotic secretion systems. Therefore, for most cases of heterologous protein secretion from yeast, it is preferable to use a yeast signal sequence and it may be simpler to do this as a matter of course with a new gene.

Use of yeast signal sequences Owing to the difficulty of predicting whether a particular foreign signal sequence will function in yeast, much work has been carried out using homologous *S.cerevisiae* signal sequences. The three most widely studied are those from acid phosphatase, invertase and α-factor.

Acid phosphatase (*PHO5*) has a classical signal sequence[7] yet there are few examples of its use for heterologous protein secretion. When used to secrete tissue-type plasminogen activator (tPA), less than 5% of the total tPA activity was found in the medium,[169] however the secretion of this protein has been problematic.[231] Human salivary α-amylase has been secreted using the *PHO5* signal[328] which in this instance gave similar results to the heterologous α-amylase and human gastrin signals and to the yeast α-factor leader.

The yeast invertase (*SUC2*) signal[286] has been used more widely for foreign protein secretion. Human α2-IFN has been secreted using this signal, which was correctly cleaved from all secreted molecules[59] unlike the native IFN signal peptide.[173] The invertase signal has also been used to secrete human α-1-antitrypsin (α-AT), however approximately 80% of the protein remained inside the cell.[262] The passage of the α-AT from the ER appeared to be the rate-limiting step in this case. Production of mouse–human chimaeric antibody has been achieved using the invertase signal to secrete immunoglobulin heavy and light chains.[180] Melnick *et al.*[255] reported the efficient secretion of human single-chain urinary plasminogen activator (scuPA) using the invertase signal.

Attachment of a heterologous protein to a yeast signal sequence will usually result in a change in the amino acid immediately C-terminal to the signal peptidase cleavage site. However, there appears to be flexibility in recognition of cleavage junctions. Fusions of the invertase signal with proteins which alter the amino acids at the junction, such as HSA or two forms of insulin, are still cleaved.[171]

The most extensively used signal sequence for heterologous protein secretion from *S.cerevisiae* is the prepro region from α-factor (*MFα1*), frequently used with the *MFα1* promoter. *MFα1* encodes a 165 amino acid protein, prepro-α-factor, which comprises a signal sequence of 19 amino acids (the pre region) and a pro region, followed by four tandem repeats of the mature 13 amino acid α-factor sequence (Figure 7). Each repeat is preceded by a short 'spacer peptide' with the structure Lys-Arg-(Glu/Asp-Ala)$_{2-3}$. Processing of prepro-α-factor

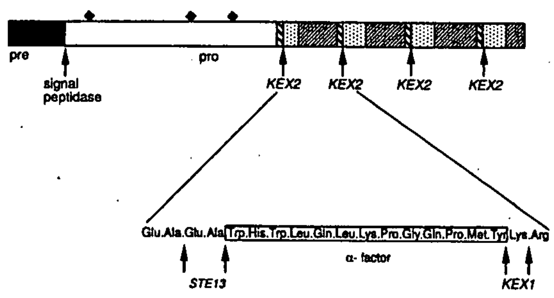


Figure 7. The structure and processing of prepro-α-factor. The product of the *MFα1* gene is shown schematically. The three N-linked glycosylation sites in the pro region are marked (◆). The peptide product from the cleavage of prepro-α-factor by KEX2 is also indicated and sites for further processing by STE13 and KEX1 are shown. See text for a description of the activities of the processing enzymes.

involves four proteolytic enzymes: (i) the pre region is cleaved by signal peptidase, (ii) the KEX2 protease cleaves on the carboxyl terminal side of the Lys-Arg sequence in the pro region and at the junction of each repeat, (iii) STE13, a dipeptidyl aminopeptidase, removes the spacer residues at the amino terminus of each repeat and (iv) the KEX1 carboxypeptidase removes the Lys and Arg residues at the carboxyl terminus of the first three repeats (see Figure 7; reviewed in reference 41).

Fusions of the *MFα1* prepro to genes encoding mature human α1-IFN,[350] a consensus α-IFN,[34] β-endorphin[34] or human EGF (hEGF)[42] resulted in efficient secretion of the heterologous protein. Processing of the prepro region took place, but the Glu-Ala spacer at the N-terminus of the secreted protein was not always removed. 5% of the hEGF and 50% of the α1-IFN was estimated to be in an intracellular form. Many foreign proteins have now been secreted from yeast using the α-factor leader, and this system has been demonstrated to be generally applicable. Changes can be made to the *MFα1* prepro to facilitate cloning, such as the introduction of a *Xho* I site towards the end of the pro region.[74] Despite a resultant amino acid change (Asp to Glu), this was shown to be fully functional.

The requirement for the pro region of the α-factor signal sequence seems to vary. Glycosylation of this region appears to be important for efficient transport of α-factor.[55] Ernst[111] reported that the pro region was not important for the secretion and processing of aminoglycoside phosphotransferase or human granulocyte-macrophage colony stimulating factor (GM-CSF). Furthermore, prepro fusions to interleukin-1β were not cleaved whereas pre fusions were correctly processed.[111] However, secretion of human insulin-like growth factor I (IGF-I) from yeast is dependent on the entire α-factor leader; the pre region is not sufficient. It may be that the pro region aids movement of the IGF-I into the ER or assists folding of the protein in the ER (A. Hinnen, personal communication). Additionally, both the pre and pro regions were required to bring about translocation of a reporter protein, α-globin.[312] Reppe *et al.*[296] compared secretion of human parathyroid hormone (hPTH) directed either by the α-factor prepro region or by the pre region alone. When the pro region was not present, there was a considerable reduction in hPTH mRNA levels, with a concomitant reduction in levels of the protein product both in the growth medium and inside the cells. This suggests that the pro region may additionally play a role in stabilizing the mRNA or may affect transcription of the gene.

Glycosylation

Many potential therapeutic proteins are glycosylated, including monoclonal antibodies, blood clotting factors, and many IFNs, hormones, growth factors and viral antigens. The carbohydrate side chains appear to be involved in diverse processes including cell–cell recognition, hormone–receptor binding, protein targetting, host-microorganism interactions, solubility and stability.[295] Glycosylation is both organism and cell-type specific and

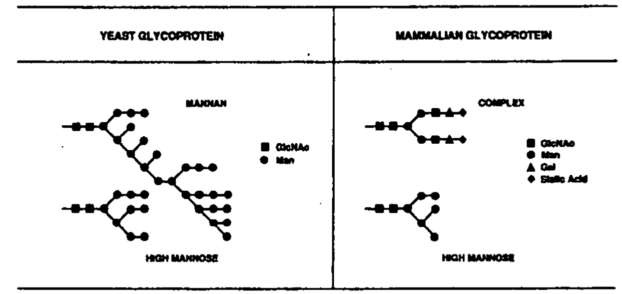


Figure 8. Comparison of yeast and mammalian glycosylation. Schematic diagram to compare the structures of oligosaccharide side chains of glycoproteins produced in yeast and mammalian cells.

therefore expression of a protein in a heterologous system will almost certainly result in a product with different modifications from the native material. This may affect the function or immunogenicity of the protein. It is interesting to note that differences in glycosylation of tPA produced in either Chinese hamster ovary or murine cell lines led to differences in the kinetics of its fibrin-dependent activation of plasminogen.[280]

Oligosaccharides may be either N-linked to asparagine or O-linked to serine or threonine residues. O-linked oligosaccharides synthesized by yeast are very different from those of higher eukaryotes, being composed only of mannose residues. Many higher eukaryotic proteins have sialylated O-linked chains of the mucin-type. N-linked glycosylation in yeast and higher eukaryotes is more conserved, and entails the addition in the ER of a core oligosaccharide unit comprising two N-acetylglucosamine (GlcNAc), nine mannose (Man) and three glucose residues. The glucose residues are subsequently trimmed from the side chain and one mannose is also removed. These steps are common to yeast, plants and higher eukaryotes.

Processing of the core oligosaccharide unit takes place in the Golgi and at this stage there is divergence between yeast and other eukaryotes (Figure 8). In animal cells further mannose residues are removed and additional sugars may be added. The resultant oligosaccharide side chains are of three types: (i) high-mannose ($GlcNAc_2Man_{2-6}$), (ii) complex, which comprises a core of $GlcNAc_2Man_3$ plus additional residues including GlcNAc, galactose (Gal), fucose and sialic acid, and (iii) hybrid, containing features of both high-mannose and complex-type oligosaccharides. (N-linked glycosylation is reviewed in reference 221). In contrast, *S.cerevisiae* does not trim back the mannose residues as it lacks the Golgi mannosidases present in higher eukaryotes. Instead, elongation of the chain may take place through a stepwise addition of further mannose residues. These additional mannose units comprise the outer chain which may be up to 75 residues long, with many branch chains. (For a review of protein glycosylation in yeast see reference 225).

Many heterologous glycoproteins have been secreted from yeast, including *Aspergillus* glucoamylase,[186] somatostatin,[148] human α-1-antitrypsin,[262] porcine urokinase,[417] human erythropoetin[106] and *Trichoderma reesei* endoglucanase I.[388] The addition of both N- and O-linked oligosaccharides has been reported (e.g. *Aspergillus* glucoamylase[186]). Analysis has revealed that not all *S. cerevisiae* N-linked oligosaccharides contain the outer chain of mannose residues. Addition of the outer chain to heterologous proteins is regarded as 'hyperglycosylation' because it results in more extensive glycosylation than is found in higher eukaryotic glycoproteins. Additionally, it should be noted that glycosylation can be very heterogeneous, giving rise to a mixed product with a varying degree of oligosaccharide attached.

Some foreign proteins are fortuitously glycosylated when secreted from yeast. This problem is more likely to occur with proteins from bacteria (for example see references 307 and 316), in which there is no glycosylation, or with proteins which are not normally secreted. Interleukin (IL)-1α and 1β are naturally-secreted proteins, but do not pass through the normal secretory pathway in mammalian cells,

.re therefore not glycosylated. However, when ted from yeast using a signal sequence, both ins are glycosylated. Although this has no sig- .nt effect on the biological activity of IL-1α,[233] :tivity of IL-1β is reduced five-fold.[234]
though many glycoproteins produced in yeast ıctive, yeast glycosylation, especially hyper- ısylation, may give rise to problems, depend- ın the intended use for the protein. In some nces, extensive glycosylation may inhibit ivity with antibodies, as was the case for EBV 0 secreted from *S.cerevisiae*.[335] This is ɔusly an undesirable effect for potential vaccine :ens. Furthermore, α1,3-linked mannose units, h occur in large numbers in the outer chain, ar to be immunogenic[16] and thus yeast-derived ›proteins may be unsuitable as therapeutics.
oblems with yeast glycosylation of foreign pro- may be circumvented by mutating the glycosy- n site(s), as was done in the case of scuPA[255] to a fully-active, non-immunogenic product. rnatively, the use of glycosylation mutants vs the production of more homogeneous pro- with limited glycosylation. Mutants in mannan ynthesis (*mnn*) may be useful for the production :terologous proteins; *mnn9* mutants do not add :xtensive outer chain of mannose units,[225] how- immunogenic terminal α1,3 mannose linkages till present. Use of the *mnn1 mnn9* strain, which tionally lacks these residues, may enable the luction of non-immunogenic, non-hyperglyco- :ed proteins. However some *mnn* mutants, e.g. 9, exhibit slower growth and greater osmotic itivity than wild-type cells. Sledziewski *et al.*[353] ted a strain in which *MNN9* expression was ɔerature-regulated, using the α2 repressor sys- in a *MAT*a *sir3*[ts] strain (see 'Transcriptional noters and terminators'). Cultures were grown igh density at the permissive temperature for *N9* expression (25°C), then switched to the non- nissive temperature (35°C). Glycoproteins pro- :d at the higher temperature were shown to have shorter, homogeneous mannose side chains :acteristic of *mnn9* mutants.
ome 'super-secreting' mutants are also hyper- osylation-deficient and may be suitable host ins for glycoprotein production. *pmr1* mutants viously called *ssc1*; see below) possess a defec- Ca^{2+} ATPase ion pump and their growth is ium-dependent.[315] The mutation is thought to se secreted proteins to by-pass the Golgi. Pro- s secreted from this strain are core-glycosylated y.

Protein folding and transport

Folding of secreted proteins occurs in the ER and involves accessory proteins such as BiP (heavy chain binding protein, the product of the *KAR2* or *GRP78* gene[272,309]), and protein disulphide isomerase.[126] Nascent proteins bind to BiP co-translocationally and are released upon folding, assembly and glycosylation; malfolded proteins bind permanently and are retained in the ER. The assembly of an oligomeric bacterial enterotoxoid in *S.cerevisiae* has recently been shown to be dependent on BiP, which acts as a molecular chaperone.[333] During the secretion of foreign proteins, problems might arise either from saturation of these accessory proteins or from their inability to aid the folding of heterologous proteins.

Glycosylation of a protein can aid both generation of the correct conformation and passage of the molecule through the secretory pathway. Inhibition of glycosylation with tunicamycin results in the intracellular accumulation of inactive invertase.[136] At lower temperatures, unglycosylated protein may be secreted: at 25°C, 50% of unglycosylated invertase is secreted.[116] Similarly, active, unglycosylated acid phosphatase can be secreted from tunicamycin-treated cells at 20°C or 25°C, but not at 30°C.[259] Unglycosylated glycoproteins are transported only if they achieve the correct conformation and the lower temperature allows this to happen. Progressive elimination, by mutation, of each of the 12 glycosylation sites of acid phosphatase resulted in increasing irreversible retention of the malfolded protein in the ER as more sites were removed.[299] At lower temperatures under-glycosylated or unglycosylated protein could be secreted.

Sato *et al.*[328] studied the expression of a series of human salivary α-amylase mutants which either lacked regions encoding cysteine residues or contained mutations at cysteine codons. The prevention of the possible formation of a disulphide bond abolished secretion. Elimination of either of the two glycosylation sites did not affect secretion and indeed α-amylase is secreted from tunicamycin-treated cells in a fully active form. Thus, at least for α-amylase, the unique conformation of the protein appears to be much more important than glycosylation for secretion in yeast. This phenomenon indicates that attempts to secrete protein fragments or fusions could lead to intracellular accumulation if normal folding and disulphide bond formation cannot take place.

In addition to requiring successful targetting and folding in the ER, a foreign protein also has to pass

through the secretory organelles in order to be released into the culture medium. Transport from the ER to the Golgi has often been shown to be rate-limiting. Hepatitis B virus large surface protein is retained in the ER, provoking the enlargement of this organelle.[31] This retention may be due to malfolding or to a specific retention signal. α-1-antitrypsin[262] and erythropoetin[106] are also retained in the ER. Soybean proglycinin, although expressed to high levels and correctly processed, was mostly insoluble and accumulated in Golgi-like structure.[386] Insolubility was due to interaction of the acidic region of the polypeptide with cellular components.[386] A Golgi or post-Golgi bottleneck was also postulated to represent a major obstacle in the secretion of IGF-I.[367] Retained material may have been malfolded and this is likely to be a common reason for intracellular accumulation of proteins within the secretory pathway.

A blockage or bottleneck in the secretory pathway caused by foreign protein secretion may also interfere with the secretion of host proteins. This was observed for acid phosphatase during the secretion of tPA[169] and may result in toxicity.

Correct folding, assembly and transport are especially important in the production of multimeric proteins. The first multimeric protein to be secreted from yeast was a mouse-human chimaeric antibody. Co-expression of immunoglobulin heavy and light chains resulted in the secretion of properly folded and assembled antibody.[180] Both whole antibody and Fab fragments were functional. The pentameric *Torpedo californica* nicotinic acetylcholine receptor with a stoichiometry $\alpha_2\beta\gamma\delta$ has also been produced in yeast by co-expression of the four subunits.[199,412] These integral membrane proteins entered the secretory pathway and were processed and glycosylated. However, no functional receptor was detected, possibly due to improper folding or assembly.

Membrane proteins often cause problems when secreted due to non-specific insertion into intracellular membranes (see 'Physiology of foreign gene expression'). EBV membrane glycoprotein gp350 was highly toxic, but could be secreted in a membrane anchor-minus form.[335] Nevertheless, some membrane proteins have been successfully produced in yeast. The human β_2-adrenergic receptor was co-expressed with a mammalian G protein.[214] Coupling of these components to each other and to downstream effectors of the yeast mating signal transduction pathway was demonstrated, indicating that correct folding and targetting took place. This *in vivo* reconstruction system provides a useful new approach for the study of signal transduction pathways.

The presence of the cell wall complicates the secretion process in yeast. Permeability may be a limiting factor and it is notable that most success has been with very small proteins. Some proteins have been reported to be localized mainly in the cell wall when secreted from *S.cerevisiae*, e.g. α-IFN[420] (166 amino acids, 20 kDa). Although it has been reported that only molecules with a molecular mass below 760 are able to diffuse freely through the cell wall, a number of very large proteins such as EBV membrane glycoprotein[335] (842 amino acids, approximately 400 kDa) and cellobiohydrolase[284] (up to 200 kDa) have been shown to be capable of passing through the cell wall and therefore this property is not related simply to the size of the molecule. It may be that there are a few large holes in the cell wall with the average pore size being small. It is difficult to draw conclusions from observations of permeability since many other factors, including strain, growth phase and composition of the medium, may have an effect on cell wall porosity (reviewed in reference 91). Additionally, glycosylation and the charge on a protein are thought to affect its passage through the cell wall. *mnn9* mutants have been reported to have increased porosity and release invertase octamers, which are normally retained.[373] *kre1* mutants, which have altered (1,6)-β-D-glucan in the cell wall, over-secrete secreted yeast proteins, possibly due to the cell wall being more leaky.[51]

Proteolytic processing

There are two ways in which proteolytic processing is relevant to heterologous protein secretion in yeast. Firstly, correct processing of the signal peptide or prepro region must take place so that the mature product is secreted. Secondly, fortuitous, undesirable processing events may occur as a result of cleavage by processing proteases.

Yeast signal peptidase (SPase) is a polypeptide complex which includes the *SEC11* gene product and a glycoprotein.[407a] Although eukaryotic SPases and the *E.coli* SPase I are disparate, it is interesting to note that homologous overproduction of the *E.coli* protein resulted in increased efficiencies of export and maturation of two poorly-processed hybrid secretory proteins.[392] This indicates that, at least in *E.coli*, the availability of the SPase can be a limiting factor. Indeed, *S.cerevisiae* strains carrying multiple copies of the wild-type *PHO5* gene

accumulate unprocessed precursor, suggesting that there is saturation of some component of the secretory pathway, possibly SPase.[155]

Deletion of the SPase cleavage site in yeast acid phosphatase leads to unprocessed, core glycosylated protein which accumulates inside the cell.[155] However, even where processing of the signal does take place, cleavage may be aberrant, giving rise to heterogeneous product. 64% of human IFN secreted from yeast using its own signal was properly processed, but 36% contained an additional three amino acids of the pre sequence.[173] Furthermore, 90% of the total IFN produced was not secreted and this intracellular material also included a third form which retained eight amino acids of pre sequence. It is unlikely that molecules which retain part or all of the pre sequence will be secreted since the hydrophobic core may be retained in the membrane.

The processing of the prepro sequence from α-factor has been described above and is a more complex process involving the *KEX2* and *STE13* gene products in addition to signal peptidase (reviewed in reference 49). The KEX2 protease cleaves on the carboxyl side of the dibasic residues, Lys-Arg and Arg-Arg. In a fusion of the α-factor prepro and a heterologous protein, these residues are at the junction, and cleavage by KEX2 liberates the heterologous protein from the leader region. The Glu/Asp-Ala spacer residues, which provide a hydrophilic environment at the KEX2 cleavage site can be dispensed with in certain cases, circumventing the problem of incomplete STE13 processing, e.g. for hEGF[42] and α1-IFN.[350] However, removal of the spacer peptide does sometimes lead to a failure in KEX2 processing, as observed for interleukin-6 (IL-6)[152] where the presence of a proline residue on the carboxyl side of the cleavage site inhibited cleavage. However, accurate recognition and cleavage occurred when the construct was modified to include alanine N-terminal to the proline. Expression of the modified gene fusion had then to be carried out in a *ste13* mutant, to avoid trimming at the N-terminus of IL-6 by the STE13 protease.

Another solution to the problem of inefficient processing is to over-express the processing enzyme genes. This approach was successfully employed in the expression of transforming growth factor α (TGFα).[19] Inclusion of the *KEX2* gene on the same multi-copy plasmid as the TGFα gene eliminated the presence of unprocessed forms of the α-factor leader-TGFα fusion protein and resulted in increased levels of secreted TGFα. Additionally, a novel *S.cerevisiae* aspartyl protease (*YAP3*) has been identified which allows KEX2-independent processing of the α-factor precursor.[102] This could be over-expressed when KEX2 processing was inefficient and limiting. Over-expression of *STE13* could also improve inefficient processing; wild-type cells carrying multiple copies of the *MFα1* gene produce mainly incompletely processed α-factor, indicating that the dipeptidyl aminopeptidase is rate-limiting.[204]

In addition to inefficient proteolytic processing, problems may also be caused by aberrant processing at internal sites in the protein. In the extreme case of β-endorphin, no complete mature protein was secreted into the medium;[34] two trypsin-like cleavage sites were observed after internal lysine residues. Use of the vacuolar protease mutant *pep4-3* did not reduce the degradation, suggesting that proteolysis was taking place during passage through the normal secretory pathway.[34] Internal processing was also a problem in the expression of hPTH in which cleavage, probably by the *KEX2* gene product, was observed after two basic residues.[131] A mutant form of hPTH which lacked one of the basic residues was no longer subject to internal proteolytic processing.[297] The mutant form retained full biological activity and was produced in significantly higher yield. Differences in proteolysis between strains and/or growth conditions may exist since two reports of secretion of GM-CSF identified different aberrant processing events. Miyajima *et al.*[257] reported cleavage after the arginine residue at position 4 of the mature protein. Price *et al.*[290] however observed both the full-length mature species and the product of cleavage after the proline at position 2 of the mature protein. Differences were also apparent in the efficiency of α-factor leader processing. This indicates that it may be possible to eliminate or to minimize internal cleavage of heterologous proteins by optimizing growth conditions and selecting the best strain.

An analysis of secreted hirudin (hir) revealed full-length hir65, but also two C-terminally degraded products, hir64 and hir63.[372] Use of the protease mutants *prc1* and *kex1* showed that this degradation was due both to yscY and yscα activities. Similarly, secreted EGF was found to be C-terminally trimmed.[74,138]

Mutants in the gene for the extracellular protease SKI5 were shown to give higher yields of secreted yeast proteins due to decreased degradation in the medium.[51]

An interesting development in protein production has been the exploitation of yeast processing

enzymes to cleave heterologous precursors. Proinsulin is converted to insulin as the result of cleavage at two dibasic sequences separated by a spacer peptide. These processing reactions can be carried out by the KEX2 protease, and various forms of proinsulin/insulin have been secreted from yeast.[376] Human proalbumin has also been shown to be processed after the Arg-Arg sequence by KEX2 to yield mature albumin.[21] *Aspergillus* glucoamylase secreted from yeast using its own signal sequence was shown to be cleaved both by SPase and by KEX2, which removed an additional six amino acids from the N-terminus.[186] This yielded a product with the same N-terminus as that of the *Aspergillus* enzyme and showed yeast to be capable of this additional processing step.

Strategies to improve secretion

Several *S.cerevisiae* strains with a 'super-secreting' phenotype have been isolated by screening for mutants with increased secretion of a particular product. Such methods have yielded strains which show a general increase in heterologous protein secretion. On finding that less than 1% of the prochymosin made in yeast was secreted, Smith *et al.*[357] employed a mutagenesis approach coupled with a rapid screening assay to isolate super-secreting colonies. The secreted material, which was activated by the low pH of the medium, was assayed by overlaying the surface of the plate with a mixture of milk and molten agarose. The chymosin clots the milk and the speed of appearance of the opaque clotted regions and their size and intensity indicate the level of prochymosin secretion. Two super-secreting strains in particular were identified and designated *ssc1* and *ssc2*.[357] The effects appeared to be additive in the double *ssc1 ssc2* mutant. The *SSC1* gene was later found to be identical to *PMR1*, which encodes a Ca^{2+} ATPase[315] (see above). The *pmr1* mutation significantly increased the secreted levels of prochymosin, bovine growth hormone and scuPA by five- to 50-fold.[357,384] However secretion of α-1-antitrypsin, which was efficiently exported in the wild-type strain, was not further improved, indicating that the *pmr1* mutation may be most effective at improving the secretion efficiency of proteins that are secreted poorly by wild-type cells.

In addition to using colony assay screens to find the most efficient secretors, Moir and Davidow[261] described a screening procedure which involved *in vitro* mutagenesis of scuPA genes on plasmids. This enabled identification of mutant forms of scuPA with either decreased or increased activity compared to the parental form. The assay uses fibrin-agar and colonies are scored by measuring zones of clearing.

Other workers have employed a similar approach to identify super-secreting strains. A plasminogen-casein assay for secreted tPA was used to isolate secretor strains;[141] previous attempts to secrete tPA from yeast had not been successful.[169,231] A general screen for secreted proteins based on a visual antibody precipitation assay has been described.[354] Repeated rounds of mutation and selection enhanced the secretion of HSA six-fold and the resultant strains were also able to produce higher levels of internally-expressed α-1-antitrypsin and human plasminogen activator inhibitor 2. This screen is suitable for any protein for which antibodies are abundantly available. Multiple rounds of mutagenesis and selection were also used to isolate a strain which secreted 70-fold more endoglucanase I than the original wild-type parent strain.[6] Again, the mutant strain was found to secrete elevated levels of other proteins.

By expressing a gene fusion of HSA and hygromycin B phosphotransferase, Chisholm *et al.*[68] selected for increased expression on the basis of the level of resistance to hygromycin B. This method was initially used to isolate strains with increased intracellular expression of the HSA fusion protein. However, when cured of the intracellular expression vector and retransformed with a HSA secretion vector, these strains also showed significant increases in secretion. Genetic analyses suggested that multiple mutations were responsible for the observed effects.

Selection may be used to identify mutants which are resistant to the toxic effects of a foreign protein. The slow growth rate of IGF-I-expressing cells was exploited by Shuster *et al.*[347] who isolated fast-growing IGF-I-resistant mutants which gradually accumulated in a population of IGF-I-expressing cells under selection. Mutations at a single locus, designated *HPX1* (for *H*eterologous *P*rotein e*X*pression) were found to confer both resistance to the toxic effects of IGF-I and its increased production.

In summary, the screening approach has frequently proved successful in the isolation of super-secreting strains, even in cases where the initial yield was high. Therefore, for industrial applications when a high yield is required, this approach deserves consideration.

Table 6. Comparison of the expression of two genes in *E.coli*, Baculovirus, *S.cerevisiae* and *P.pastoris*.

Expression system	Expression level (%) Fragment C	Pertactin
E.coli	24[241a]	30[242a]
Baculovirus	10‡,[60a]	>40*
S.cerevisiae GAL7	2–3[307]	0·1[306]
P.pastoris	28§,[73]	10¶,[306]

‡Estimate; yield variable and greatly reduced on scale-up.
*Estimate; yield variable and greatly reduced on scale-up; I.G. Charles, personal communication.
§Scale-up without loss of yield to >12 g/l.
¶Scale-up without loss of yield to >4 g/l.

EXPRESSION IN NON-*SACCHAROMYCES* YEASTS

Introduction

In some ways *S.cerevisiae* could be regarded as a non-optimal host for the large-scale production of foreign proteins due to drawbacks such as the lack of very strong, tightly-regulated promoters, the need for fed-batch fermentation to attain high-cell densities, and hyperglycosylation. Although these have been addressed by exploiting its sophisticated molecular genetics, another important approach has been to develop expression systems in other yeasts. Most of these alternative systems are based on commercially-important yeasts that have been selected for their favourable growth characteristics at industrial scale, or on yeasts which have other favourable intrinsic properties (e.g. high-level secretion).

The most extensively developed system is based on *Pichia pastoris*. An efficient and tightly-regulated promoter coupled with very straightforward techniques for high-biomass cultivation make this a powerful expression system. It is currently the simplest of eukaryotic systems to scale up, and there are now several comparative studies suggesting it can be used to avoid limitations on transcription which are sometimes encountered with *S.cerevisiae* (see Table 6). *Hansenula polymorpha* has similar properties and there are some promising examples of its use. A potential problem however, especially with toxic products, may be that under the conditions normally used for high-density growth, expression is significantly derepressed. *K.lactis* and *Y.lipolytica* are both industrial yeasts which have been examined largely because of their capacity for high-level secretion. *K.lactis* is similar to *S.cerevisiae* and has well-developed molecular genetics. Unlike the other systems discussed here it has the advantage of highly stable episomal vectors. Finally the distantly-related fission yeast, *S.pombe*, has very advanced genetics but its use in foreign gene expression has mainly been in isolating and studying homologous mammalian genes, rather than protein production. The following sections will briefly review these non-*Saccharomyces* heterologous expression systems. A summary of the various proteins that have been expressed in each system is given in Table 7.

Pichia pastoris

The methylotroph *P.pastoris* has now been used to express high levels of many different intracellular and secreted proteins. Strains were originally developed for the large-scale, high-yield production of single-cell protein using defined medium containing methanol.[401] This fermentation technology subsequently provided the basis for an efficient expression system which uses the promoter from the tightly-regulated *AOX1* gene.[81] *AOX1* encodes alcohol oxidase, which catalyses the first step in the assimilation of methanol, and can be induced to give levels of up to 30% of t.c.p. by the addition of methanol.[76]

Since *P.pastoris* has no native plasmids (other than linear DNAs, K. Sreekrishna, personal communication) expression vectors designed for chromosomal integration have been developed. These vectors have used the *HIS4* gene for selection, but a limited number of alternative selection markers are now available, e.g. *ARG4* from *S.cerevisiae*[80] or *P.pastoris* (J. Cregg, personal communication), the *Tn903* G418-resistance gene (K. Sreekrishna personal communication), and the *S.cerevisiae SUC2* gene[361] which can be used as a dominant marker since *P.pastoris* cannot use sucrose as the sole carbon source. Multi-copy episomal vectors containing the *P.pastoris* ARS sequences, PARS1 and PARS2, have also been constructed.[79] However, for maximum stability, particularly in large-scale inductions, the expression cassette has been transplaced into *AOX1*, or integrated by single cross-over into *AOX1* or *HIS4* (see Figure 9). It is worth noting that, using *HIS4* selection, *HIS4* integrants are the least stable type of transformant since recombination can result in excision of the expression cassette leaving behind a wild-type marker. Although transplacement into

AOX1 results in gene disruption, transformants are still capable of slow growth on methanol due to the presence of a second alcohol oxidase gene, *AOX2*, which is less well expressed.[80]

The unusually high level of alcohol oxidase together with evidence for transcriptional regulation[107] suggested a very powerful and efficiently regulated promoter suitable for foreign gene expression. This was first tested by isolating the *AOX1* promoter, and also that from the co-regulated dihydroxyacetone synthase gene (*DHAS*), and fusing them to the *lacZ* gene.[380] In single-copy the *AOX1-lacZ* hybrid gene was tightly repressed by glucose or glycerol and was efficiently induced by methanol; with multi-copy *PARS* vectors some expression in glycerol occurred. The *AOX1* promoter has subsequently been used to produce a variety of foreign proteins. Promoters from several other *P.pastoris* genes have been isolated, e.g. from the constitutive *GAP* and *PGK* genes, and from the *AOX2* gene,[80] but there are no published examples of their use.

A number of factors may potentially affect expression levels using *P.pastoris* integration vectors. In particular, disruption of *AOX1* by transplacement might be expected to have important physiological consequences which may influence foreign gene expression. During induction *aox1* transformants do not simultaneously produce high levels of alcohol oxidase and heterologous protein, and they also grow more slowly ('*m*ethanol-*u*tilization *s*low', Mut^s), and have a much lower O_2 demand than wild-type (Mut^+) strains. Indeed, in the first report of foreign gene expression in *P.pastoris*, a higher proportion of immunogenic HBsAg particles were produced in a Mut^s host compared to Mut^+, although expression levels were similar.[81] This was attributed to one or more events during assembly being rate-limiting in fast-growing cells, though there is evidence that particle maturation in *S.cerevisiae* occurs during protein extraction.[399] The parameters affecting foreign gene expression in *P.pastoris* were investigated in a study using tetanus toxin fragment C.[73] In a direct comparison, similar amounts were produced in Mut^S and Mut^+ hosts, even at levels approaching 28% t.c.p. Expression levels were also independent of the site (*HIS4 versus AOX1*) and type (single *versus* double cross-over) of integration. However, in another study, a Mut^S integrant expressed *Bordetella pertussis* pertactin at 2% t.c.p. in shake flasks and 10% t.c.p. in fermenters, while a Mut^+ strain gave 5–6% t.c.p. in shake-flasks or fermenters.[306] Thus, the significant physiological differences between Mut^S and Mut^+ strains during induction may in some instances influence the expression of foreign genes.

The abundance of alcohol oxidase led to the assumption that single-copy integrants would yield sufficient levels of foreign proteins using the *AOX1* promoter. However, with single-copy vectors it has not been possible to obtain yields of heterologous proteins that are as high as alcohol oxidase.[73,74,306] This is not due to the presence of DAS elements since, unlike the *S.cerevisiae PGK* gene[253], replacement of the *AOX1* coding region with foreign sequences does not affect mRNA levels[73] (R. Buckholz personal communication). Rather, higher levels of alcohol oxidase are probably due to its exceptional proteolytic stability which may be enhanced as a result of sequestration to peroxisomes. Multi-copy integrant strains are now used for increased yields of foreign proteins. These can be obtained using vectors containing multiple expression cassettes[375] or, for very high copy number, by screening for multiple integration events.[73,74,306] Surprisingly, very high copy number integrants can be isolated from transformations using DNA fragments designed for single-copy transplacement. Sreekrishna *et al.*[359] observed a high degree of variation (from 1–30% t.c.p) in the level of tumour necrosis factor (TNF) expressed in different Mut^S transformants and found this was due to differences in gene copy number.

The mechanism of this 'multi-copy transplacement' was initially unclear and it was not apparent whether this was a general phenomenon, nor whether high-level expression would invariably occur as a result. However, multi-copy transplacement has now been observed in several other instances.[73,74,306] The use of such strains has frequently resulted in remarkably high yields which compare very favourably with other expression systems (see Table 6). These studies also show that multi-copy transplacement strains are stable during high-density growth and induction. The clonal variation in vector copy number and protein yield among transplacement transformants expressing pertactin is illustrated in Figure 10.

Multi-copy transplacement occurs at a variable frequency, and is observed in about 1–10% of Mut^S transformants.[73,74,306] To understand the mechanism involved and the factors controlling this, Clare *et al.*[73] carried out a detailed DNA analysis of fragment C-vector transformants. An interesting finding was that 75–95% of transformants had not undergone a transplacement event (i.e. were Mut^+) but were single or multiple integrations at *AOX1* or

Table 7. Production of foreign proteins in non-*Saccharomyces* yeasts.

Yeast	Protein[a]	Location[b]	Promoter[c]	Reference
Pichia pastoris	β-galactosidase	I	AOX1, DHAS	380
	HBsAg	I	AOX1	81
	Tetanus toxin fragment C	I	AOX1	73
	Pertactin	I	AOX1	306
	TNF	I	AOX1	359
	Streptokinase	I	AOX1	154a
	SOD	I	AOX1	43, 375
	HIVgp120	I, S	AOX1	C.A.S., R. Buckholz, unpublished results
	S.c. invertase	S	AOX1	381
	Bovine lysozyme	S	AOX1	97
	Human EGF	S	AOX1	43
	Murine EGF	S	AOX1	74
	Aprotinin	S	AOX1	375
	HSA	S	AOX1	K. Sreekrishna, personal communication
Hansenula polymorpha	β-lactamase	I, S	MOX, FMD, DAS	198
	HBsAg	PERI	MOX, FMD	340, 197
	PreS1-S2-HBsAg	PERI	MOX	197
	α-galactosidase	S	MOX	115
	Glucoamylase	S	FMD	137
	HSA	S	MOX	174
	S.c. invertase	S	MOX	198
Kluyveromyces lactis	Prochymosin	S	LAC4	391
	IL-1β	S	*S.c.*PHO5,*S.c.*PGK	119
	HSA	S	LAC4,*S.c.*PHO5,*S.c.*PGK	120
	HSA-CD4	S	*S.c.*PGK	R. Fleer, personal communication
	Schw.o. α-amylase	S	Homologous	368
	tPA	S	?	411
	TIMP	S	?	411
Yarrowia lipolytica	β-galactosidase	I	LEU2	134
	S.c. invertase	S	XPR2	270
	Bovine prochymosin	S	XPR2, LEU2	124
	Porcine IFN	S	XPR2	164, 271
Schizosaccharomyces pombe	Polyoma middle-T Ag	I	*S.c.*PGK	27
	β-galactosidase	I	54/1, fbp, adh, GRE, CaMV35S*	224, 176, 287
	CAT	I	nmt1, HCGα, CMV*, SV40*,GRE	249, 379, 287
	Human epoxide hydrolase	I	adh	192
	Factor XIIIa	I	adh	45
	IBD virus VP3	I	adh,*S.c.*ADH1	194
	E.coli β-glucuronidase	I	CaMV35S*	289
	Single-chain Ab	I	adh	88
	Bacterio-opsin	PLM	adh	166
	STP1 glucose transporter	PLM	adh	329
	S.c. invertase	PERI	Homologous	263
	S.dia. glucoamylase	PERI?	Homologous	112
	S.c. α-mannosidase	CWALL?	Homologous	226
	S.c. exoglucanase	CWALL?	Homologous	226
	S.c. endochitinase	CWALL?	Homologous	226
	Antithrombin III	S	*S.c.*ADH1,*S.c.*CYC1	46
	Schw.o. α-amylase	S	Homologous	368[a]

[a] *S.c.*, *Saccharomyces cerevisiae*; *Schw.o.*, *Schwanniomyces occidentalis*; *S.dia.*, *Saccharomyces diastaticus*.
[b] Location of expressed protein: I, intracellular; PERI, periplasmic; PLM, plasmamembrane; CWALL, cell wall; S, secreted.
[c] Promoters given are native to the organism except: *S.c.*, *Saccharomyces cerevisiae*; *, viral; homologous, homologous to the gene expressed; GRE, glucocorticoid response elements; HCGα, human chorionic gonadotrophin α.

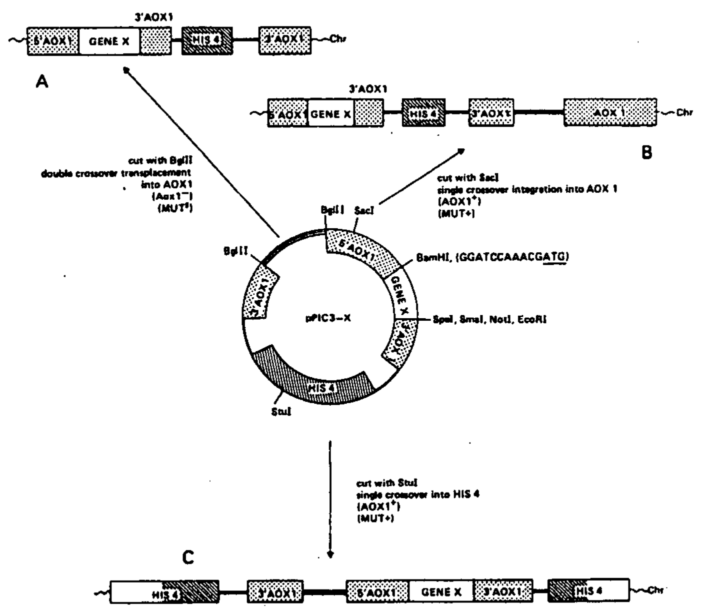


Figure 9. Plasmid map of the *P.pastoris* expression vector pPIC3 showing the three different modes of chromosomal integration. (A) Digestion with *BglII* yields a fragment with ends that are homologous to the 5' and 3' regions of *AOX1* that is targetted to transplace into *AOX1*. (B) Linearizing with *SacI* targets integration by single cross-over into *AOX1*. (C) Linearizing with *StuI* targets single cross-over integration into *HIS4*.

HIS4, or lacked vector sequences and were presumably *HIS4* gene convertants. The multi-copy transformants predominantly contained head-to-tail tandem arrays of the transplacing fragment located at *AOX1*. Both these and the other structures observed could be explained by intramolecular ligation of transplacement cassettes *in vivo*, prior to repeated single cross-over integration into *AOX1*, *HIS4* or into previously transplaced vector. Intermolecular ligation also occurred at low frequency since integrants containing the entire vector or head-to-head repeats were also found.

This mechanism of multi-copy formation (i.e. repeated single cross-overs of circularized transplacement cassettes) suggests that DNA fragments designed for targetted single cross-over integration rather than transplacement could be used to generate multi-copy transformants more readily, since higher transformation frequencies can be achieved. For the routine isolation of high-copy integrants, mass screening methods based on colony hybridization have been used[74,306] (Figure 10). More recently, vectors containing the G418-resistance marker have been used to identify high-copy integrants by resistance to increasing concentrations of the drug (K. Sreekrishna, C.A.S., unpublished results). Thus, the most efficient procedure for isolating high-copy integrants is probably to use these vectors in single-cross-over integrations and to select clones, from amongst primary His$^+$ transformants, that are resistant to high concentrations of G418 (e.g. 2 mg/ml). Multi-copy Mut^S clones can be obtained by using a strain carrying a disrupted copy of *AOX1* (e.g. KM71[80]).

With some foreign proteins there is a direct correlation between gene dosage and expression level, although the yield per expression unit is usually reduced at very high copy number. This can be clearly

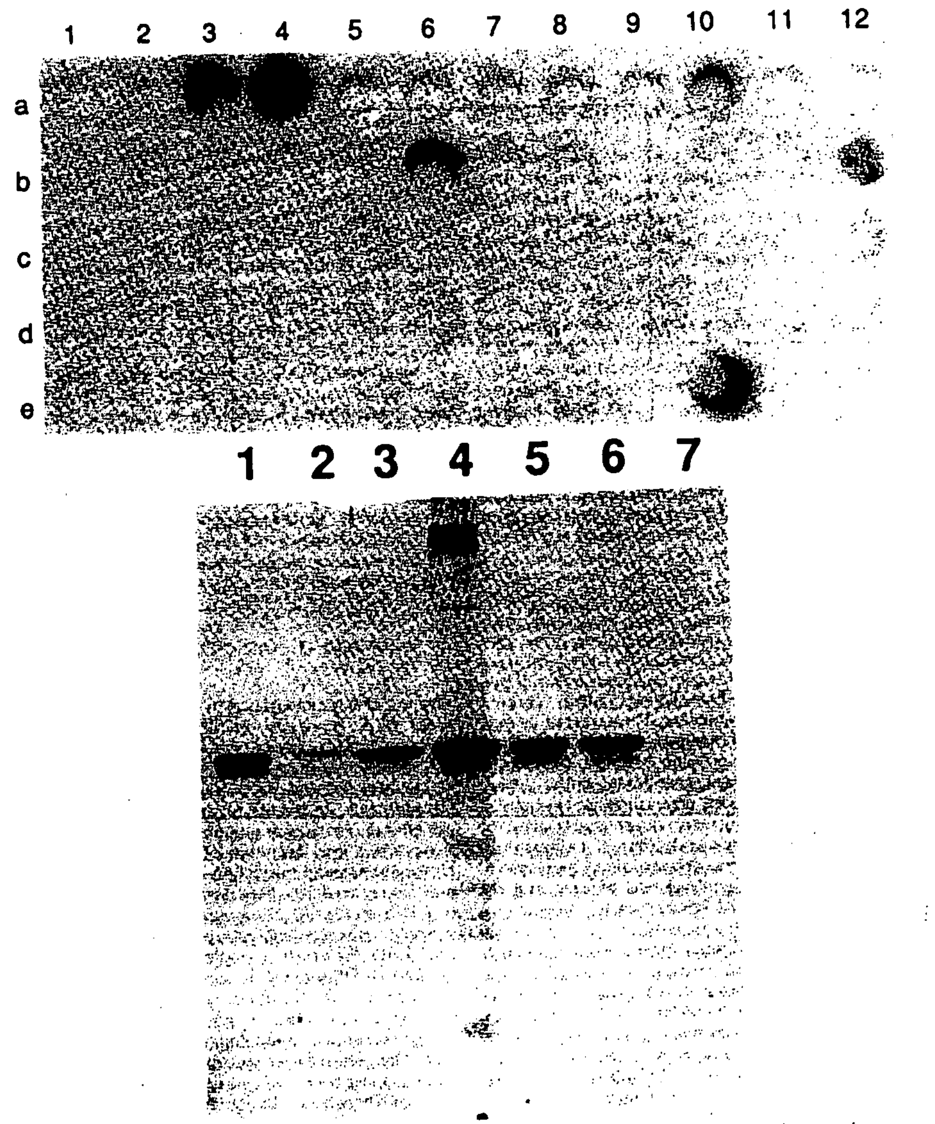


Figure 10. Screening for high copy number *Pichia* integrants. Mut$^+$ transformants of a pPIC3 vector expressing pertactin were isolated prior to screening for high copy number.[306] (A) DNA dot blot of intact whole cells which were grown in microtitre wells then lysed on nitrocellulose. Most dots correspond to single-copy transformants (e.g. a1) while some are multi-copy (e.g. a4). (B) Western blot showing variation of pertactin expression levels among transformants showing a correlation with vector copy number. Track (1) native pertactin, (2) single-copy *Pichia* transformant, (3) 13-copy, (4) 30-copy, (5) 12-copy, (6) 21-copy, (7) *S.cerevisiae* pWYG7-based vector.

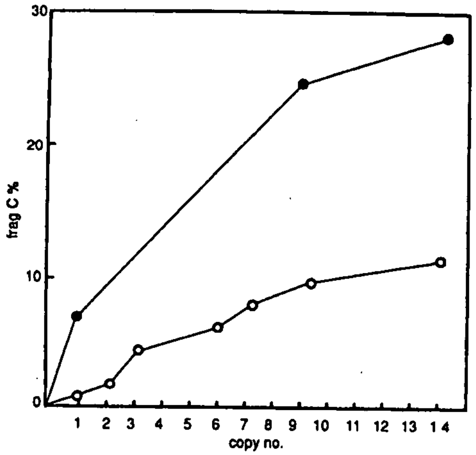


Figure 11. The correlation between gene dosage and the level of expression of tetanus toxin fragment C in Mut^s *P.pastoris* transformants. Closed circles correspond to high density inductions in the fermenter, open circles represent shake-flask inductions.

seen in Figure 11, which shows fragment C expression at varying copy number. The nature of the gene and its protein product may determine the gene dosage at which expression becomes limited. For toxic proteins increasing gene dosage will not necessarily improve yields, e.g. increasing the copy number from one to three did not result in higher levels of the secreted enzyme bovine lysozyme, although the amount of mRNA was proportionally increased.[375] Indeed, the secretion level was reduced and the enzyme accumulated intracellularly, suggesting a blockage of the secretory pathway.

In addition to efficient expression of intracellular products, *P.pastoris* is also capable of secreting high levels of foreign proteins. This was first demonstrated using the *S.cerevisiae SUC2* gene product, invertase, which accumulated to 2·5 g/l in the medium.[381] The *P.pastoris*-secreted product had much shorter N-linked carbohydrate side chains than that of *S.cerevisiae* (8–14 mannose units compared to >50). In addition, these did not contain the terminal α1,3-mannose linkages[379a] which are present in *S.cerevisiae* glycoproteins and have been shown to be immunogenic.[18] Studies of bulk endogenous glycoproteins from *P.pastoris* revealed that they are also mannosylated to a lesser extent than those of *S. cerevisiae*—about 35% of N-linked oligosaccharides having less than 14 mannose units.[149] This suggests that *P.pastoris* may be a better host for the production of heterologous glycoproteins, since extensive outer chain structures can alter the enzymic or immunogenic properties of the product (see 'Secretion of foreign proteins'). However, hyperglycosylation of certain foreign proteins secreted from *P.pastoris* has been observed (e.g. HIV gp120, C.A.S., unpublished results).

Several other proteins have been secreted at high level from *P.pastoris*, including bovine lysozyme (0·55 g/l[97]), HSA (3 g/l, K. Sreekrishna, personal communication), aprotinin (0·93 g/l[375]), hEGF (0·4 g/l[43]) and mEGF (0·45 g/l[74]). In some of these examples the corresponding signal peptide was used. The prepro leader sequence of *S.cerevisiae* α-factor has also been used, and is efficiently processed by a *KEX2* protease-like activity in *P.pastoris*.[74,375] Interestingly, Thill *et al.*[375] have reported that *S. cerevisiae* invertase is secreted by *Pichia* with much faster kinetics using the α-factor prepro sequence than with its own secretion signal. This may be an important factor in determining the gene dosage at

which secretion of a particular protein becomes limited and may also influence the glycosylation pattern of the final product (J. Tschopp, personal communication).

A problem that has occurred with several proteins secreted by *P.pastoris* is proteolytic instability in the culture medium, e.g. mEGF,[74] hEGF,[43] HSA (K. Sreekrishna, personal communication), HIV gp120 (C.A.S., unpublished results). This can be minimized by altering the pH of the culture medium upon induction (e.g. raising to pH 6·0,[74] or lowering to pH 3·0, G. Thill, personal communication), or by using a strain deleted in the *PEP4* homologue (M. Gleeson, personal communication). In the long term it would be desirable to identify the specific extracellular proteases responsible in order to construct strains lacking these activities.

An advantage of the *Pichia* system is the ease of high-density growth and scale-up without any reduction of specific productivity.[81] This feature is especially advantageous for secreted products since the concentration in the medium increases with cell density. In fact shake-flask inductions are normally sub-optimal, and there is often a large improvement in productivity using controlled fermenters[73,74,306] probably due to the high O_2 demand of the organism.

Hansenula polymorpha

The expression system developed in the methylotroph, *H.polymorpha*, is similar to that of *P.pastoris*. The gene encoding the peroxisomally-located enzyme, methanol oxidase (*MOX*), has been isolated[229] and the promoter used to express foreign genes. As with *P.pastoris AOX1*, the *MOX* gene is highly-expressed, giving methanol oxidase levels up to 37% t.c.p.,[142] and its transcription is tightly regulated. One important difference, however, is that expression of the *H.polymorpha* gene is significantly derepressed during glucose limitation or in the absence glucose, e.g. using substrates such as glycerol, sorbitol or ribose.[103,104] Thus, tight regulation of the promoter is lost in the conditions normally used for high-biomass fermentations.[137]

Transformation systems were developed using the *LEU2* and *URA3* genes from *S.cerevisiae* as selectable markers. The *LEU2* vector, YEp13, gave a low frequency of transformants which maintained the plasmid autonomously, but with low mitotic stability.[143] No integration was observed, even with linearized vector DNA, suggesting a limited homology between *H.polymorpha* and *S.cerevisiae* *LEU2* genes, although integration has subsequently been shown to occur.[348] The bacterial plasmid sequences may promote autonomous replication since the 2μ replication origin probably does not function in *H.polymorpha*.[302] The *URA3* vector, YIp5, gave a low frequency of very unstable transformants that occasionally gave rise to stable derivatives in which the plasmid had randomly integrated.[302] The transformation frequency was significantly increased by the presence of two independently isolated chromosomal DNA sequences, *HARS1* and *HARS2*, which also conferred unstable autonomous replication. Following prolonged propagation of such transformants in non-selective medium, stable *Ura*+ segregants could be isolated which contained multiple (up to 75), tandemly integrated copies of the vector. Although the mechanism of multi-copy integration is unclear, this procedure can be used to generate stable, high-copy strains for the expression of foreign proteins.[197] The use of the dominant G418-resistance marker in *H.polymorpha* has also been described.[144,197]

The *H.polymorpha* system has been used for the efficient expression and assembly of HBsAg particles containing preS2 sequences.[340] More than 95% of the protein produced could be recovered as particles, over half of which were secreted to the periplasm and could be released into the medium by inducing cells in the presence of β-glucanase. The overall yield in these permeabilized cells was several fold higher than in untreated cells, suggesting that the intracellular accumulation of particles inhibited further synthesis. Janowicz *et al.*[197] produced mixed particles containing preS1-S2-HBsAg and HBsAg by co-expression with two promoters (from the methanol-regulated formate dehydrogenase gene, *FMD*, and from *MOX*) and by selection with *URA3* and G418. Multiple integration of both genes was achieved using HARS vectors by the method described above. By obtaining strains with different copy numbers of each gene, composite particles containing various ratios of preS1-S2 and S antigen could be produced, with a total yield of 2–8% t.c.p.

The secreted plant enzyme, α-galactosidase has been produced in *H.polymorpha*.[113] The *S.cerevisiae* invertase signal sequence was used, which gave efficient secretion and was correctly processed. Attempts to isolate high-copy integrants with the YEp13-derived expression vector, by passage of transformants in non-selective medium, resulted in a strain containing two copies integrated at *MOX* and one at *LEU2*. This low integration frequency

and copy number may be because 2μ sequences prevent integration or are unstable when integrated,[348] or because the plasmid is less stable than HARS vectors and is lost before multiple integration can occur. The three-copy strain secreted 42 mg/l, which was equivalent to about 5% t.c.p. The *H. polymorpha*-derived material was over-glycosylated compared to the native enzyme and had a lower specific activity. Full activity could be restored on treatment with endoglycosidase H.

As with *P.pastoris*, *H.polymorpha* can be grown to high density (100–130 g/l), resulting in very high volumetric yields of secreted proteins. Using an integrant containing four copies of a *Schwanniomyces occidentalis* glucoamylase gene, up to 1·4 g/l of secreted enzyme was obtained.[137] The yield from an eight-copy integrant was much lower suggesting a blockage of the secretory pathway at the higher gene dosage. The secretion of several other foreign proteins in *H.polymorpha* has been reported including HSA[174], invertase and β-lactamase.[198]

Kluyveromyces lactis

K.lactis has been used in the food industry for many years in the production of β-galactosidase (lactase). Thus, its large scale cultivation has been extensively studied, and it is well accepted for the production of proteins for human use. The ability to grow on cheap substrates, such as lactose and whey, further increases its potential as a host for the production of heterologous proteins, especially for low-value products.

Transformation systems were initially developed by isolating *K.lactis* ARS sequences, since neither the *S.cerevisiae ARS1* nor 2μ replicates in *K.lactis*.[84,362] However, as with *S.cerevisiae*, *K.lactis* ARS vectors are highly unstable and are of limited use in expression systems. A number of selection markers are available for *K.lactis*, e.g. *S.cerevisiae TRP1*[84] and *URA3*,[90] *K.lactis TRP1*,[365] *URA3*[346] and *LAC4*,[84] and the G418-resistance gene.[84,362]

The two cytoplasmic linear plasmids, k1 (8·9 kb) and k2 (13·4 kb), present in killer strains of *K.lactis* have been considered as a potential vector system. They are stably maintained at 100–200 copies per cell and the regions of k1 that encode killer toxin can be deleted without affecting maintenance (reviewed in reference 364). However, their 5'-termini are covalently linked to protein, hindering manipulation *in vitro* and amplification in *E.coli*. Additionally, these plasmids encode their own cytoplasmic transcription system which does not recognize nuclear promoters. The difficulties in manipulation can be overcome by targetted integration of foreign DNA into native k1. If conventional nuclear selection markers are used, e.g. *LEU2*, this results in linear nuclear plasmids containing telomeres.[206] However, by fusing such markers to k1 promoters, recombinant linear plasmids which are cytoplasmic and stable can be generated.[207,374] It should be possible to use this system for foreign gene expression, although the k1/k2 promoters appear to be rather weak and further development, e.g. the use of the bacteriophage T7 transcription system, may be necessary.

Stable high-copy *K.lactis* expression vectors have been constructed based on the *Kluyveromyces drosophilarum* plasmid, pKD1. Although there is little sequence similarity, pKD1 is organizationally very similar to the 2μ plasmid of *S.cerevisiae*. It encodes analogous replication, amplification and segregation functions,[65] which are also active in *K.lactis*.[30] Several different types of vector based on pKD1 have been constructed which behave similarly to the analogous 2μ vectors.[30] Vectors carrying just the *cis*-acting replication element, located near one of the inverted repeats, can be maintained in host strains which have resident pKD1. However, vectors containing the entire pKD1 plasmid are significantly more stable and can be maintained in any *K.lactis* host strain.[64] The unique *Eco*R1 site adjacent to one of the inverted repeats can be used to insert foreign DNA without interruption of plasmid functions.[64] Such vectors are highly stable in pKD1° strains, even in the absence of selection, although stability is somewhat reduced in pKD1$^+$ hosts, perhaps due to incompatibility or competition. These vectors are currently used for optimal foreign gene expression in *K.lactis*.[119,120] Another reported method of producing highly stable pKD1-based vectors is to transform pKD1$^+$ hosts with plasmids containing one of the inverted repeats, in addition to the desired foreign DNA. Once introduced, these can recombine with resident pKD1 at relatively high frequency, due to the pKD1-encoded FLP recombinase, giving stable recombinant vectors.[402]

A small number of promoters have been used in *K.lactis* expression vectors. The best-characterized *K.lactis* promoter is that of the *LAC4* gene, encoding β-galactosidase, which is induced up to 100-fold by lactose or galactose. Its regulation parallels the *S.cerevisiae GAL* system, though there are significant differences: the *K.lactis GAL* genes, including *LAC4*, show no glucose repression, and the *GAL1,7* and *10* genes are only induced five- to ten-fold.[96a] Some of these differences can be ascribed

to the fact that *LAC9*, the *GAL4* homologue of *K.lactis*, is not involved in glucose repression in many strains. Several *S.cerevisiae* promoters are active in *K.lactis* and have been used for foreign gene expression, e.g. *PGK*[120] and *PHO5*.[119] In future, promoters from the recently isolated *K.lactis* genes encoding GAP[344] and alcohol dehydrogenase[324,344] may be used.

A number of studies show that *K.lactis* can efficiently secrete foreign proteins, including prochymosin which is only poorly secreted by *S.cerevisiae*. Signal peptides derived from HSA, the *K.lactis* killer toxin α-subunit, and the prepro peptides of α-factor from both *K.lactis* and *S.cerevisiae* have been used. In marked contrast to *S.cerevisiae*, prochymosin was efficiently secreted by *K.lactis* in a fully soluble, acid-activatable form, using a single-copy integration vector carrying the *LAC4* promoter.[391] Remarkably, about 80% of the prochymosin produced was secreted even when expressed without a signal peptide, though the overall level was reduced. The highest levels were obtained using the native leader peptide or the α-factor prepro sequences from either *K.lactis* or *S.cerevisiae*. The stability of the integrated expression cassette was sufficient to allow scale-up to 41,000 litres, and the product is used commercially in the manufacture of milk products.

Other heterologous proteins have been secreted using pKD1-derived vectors. Fleer *et al.*[120] described the secretion of HSA using its own secretion signal, or that of the *K.lactis* killer toxin α-subunit. The level of HSA produced was highly strain-dependent and this was largely due to differences in stability of the vector. Using the *S.cerevisiae PGK* promoter, the highest-expressing strain produced about 300 mg/l HSA in shake flasks. In high density (80–90 g/l dry weight of cells), fed-batch fermentations several g/l HSA were produced from cultures of up to 1000 litres. This system has also been used for the production of a HSA-CD4 fusion protein designed as a potential therapeutic agent for use in HIV infection (R. Fleer, personal communication).

The secretion of interleukin 1β (IL-1β) by *K.lactis* has also been reported, using the toxin α-subunit signal, either with or without the pro region derived from HSA.[119] These signal peptides were accurately and efficiently processed, but only about 20% of the product was secreted (80 mg/l in shake flasks). As in *S.cerevisiae*, the *K.lactis*-derived IL-1β was fortuitously glycosylated and largely inactive, though full activity was restored by digestion with endoglycosidase H. A mutant form of IL-1β which lacked the unique N-linked glycosylation site was also fully active. Plasmid stability was significantly reduced in IL-1β expressing cells, but this could be improved by replacing the constitutive promoter (*S.cerevisiae PGK*) with a regulated one (*S.cerevisiae PHO5*).

The glycosylation pattern of two other heterologous proteins secreted in *K.lactis* was examined by Yeh *et al.*[411] Variants of tPA and the tissue specific inhibitor of metalloproteinases (TIMP) were secreted using their own secretion signal or that of the killer toxin. In each case the secreted material was reported to be over-glycosylated and was not immunoreactive unless treated with endo H.

Yarrowia lipolytica

Y.lipolytica has been investigated for use in a number of industrial processes, including the production of a various metabolites (e.g. citric acid,[337] 2-keto glutarate,[243] erythritol,[93] mannitol,[94] isopropyl malate[92]), the bioconversion of alkanes and fatty acids into alcohol, and the production of single cell protein from *n*-paraffins.[133,163] In addition, this yeast secretes a variety of high molecular weight proteins including acid proteases,[408] lipases,[2] a ribonuclease,[66] and an alkaline extracellular protease (AEP).[377] Under optimal conditions AEP can be induced to levels of 1–2 g/l and is the major component of the culture supernatant.[377] This inherent capacity for high-level secretion, plus the ability to grow to high cell density at industrial scale, have prompted the investigation of *Y.lipolytica* as a host for heterologous gene expression.

Early studies of *Y.lipolytica* genetics were hampered by its low sporulation frequency and low spore viability. Indeed, these characteristics led to its original classification as *Candida lipolytica*, since no sexual cycle was demonstrated until 1970.[403] However, improved techniques for mating, sporulation, and ascospore recovery, combined with improved strains derived from inbreeding programmes have led to the development of a genetic map, comprising at least five linkage groups.[273] *Y.lipolytica* is a dimorphic yeast, being unicellular in minimal medium containing glucose or *n*-hexadecane, forming mycelia in minimal medium containing olive oil or casein, and giving a mixture of both forms in complex medium. Mutants which form smooth colonies containing only yeast-phase cells have been isolated[123,274] but the molecular events involved in the regulation of growth morphology are uncharacterized.

Transformation of *Y.lipolytica* was first achieved using two different selectable markers. Davidow

et al.[85] used the homologous *LEU2* gene to transform cells permeabilized by lithium acetate. Transformation occurred by integration into the chromosomal *LEU2* locus via homologous recombination. Linearization of the vector within the *LEU2* gene resulted in a 1000-fold increase in transformation frequency. Gaillardin *et al.*[135] used vectors containing random *Y.lipolytica* genomic fragments inserted into the upstream region of the *S.cerevisiae LYS2* gene for the selection of transformed sphaeroplasts. Transformants commonly contained several tandemly integrated copies of the vector suggesting that the *LYS2* gene was poorly expressed. Unexpectedly, no sequences capable of supporting autonomous replication were isolated. Wing and Ogrydziak[405] were also unable to isolate ARS sequences using a similar strategy. This was subsequently found to be because ARSs occur relatively infrequently in *Y.lipolytica*, and because plasmids containing these sequences are unusually stable.[123] Thus, such transformants were probably erroneously scored as integrants. Using mutant host strains in which replicating vectors were fortuitously less stable, Fournier *et al.*[123] isolated the *Y.lipolytica* ARS elements, *ARS18* and *ARS68*. Vectors containing these sequences were present in only 1–3 copies per cell and were very stable, being lost at a rate of 0·5–5% per generation. Since only two different ARSs were isolated, a maximum frequency of only one element per 1000 kb of genomic DNA can be calculated. These characteristics suggested that the *ARS18* and *ARS68* elements could also contain centromere function, and this has been demonstrated genetically.[133] Expression vectors based on the *ARS18* element have now been developed.[271]

In addition to *LEU2*, several other *Y.lipolytica* genes have now been cloned and used as selection markers for transformation, e.g. *LYS1*, *LYS5* and *ADE1*,[163,407] *HIS1* and *URA3*.[86] *Y.lipolytica* is resistant to many of the antibiotics commonly used for *S.cerevisiae*, including G418 and chloramphenicol, but alternative dominant selection markers have been developed. The *LEU2* promoter was fused to the phleomycin-resistance gene from Tn5 and used for the direct selection of transformants.[134] A high proportion of resistant colonies were found to be untransformed mutants, but this could be reduced by introducing an expression phase prior to plating. An alternative dominant transformation marker is the *SUC2* gene of *S.cerevisiae*, which was fused to the promoter and secretion signal sequence from the AEP gene, *XPR2*.[270] Direct selection on sucrose medium was possible, probably since most of the secreted invertase remains in the periplasm, preventing crossfeeding of untransformed cells.

The *XPR2* promoter, which is tightly regulated by pH and by carbon and nitrogen sources,[275,276] and the *XPR2* secretion signal sequence have been used to produce other foreign proteins in addition to invertase. The *XPR2* gene was independently cloned by three groups and found to encode a preproenzyme.[87,247,270] The mature protease is produced by a series of processing events which sequentially remove the presequence and two or three proregions.[247] The presequence is cleaved on translocation into the endoplasmic reticulum and, after trimming of N-terminal X-Pro and X-Ala dipeptides, the pro sequences are then cleaved by a KEX2-like activity, most likely encoded by the *XPR6* gene. Deletion analysis suggests that the proregion is required for efficient secretion of AEP.[133]

The AEP prepro sequences have been used to direct the secretion of two commercially-important proteins, bovine prochymosin[124] and porcine α-interferon.[164,271] Franke *et al.*[124] fused prochymosin cDNA to five positions within the AEP coding sequence, resulting in the expression of prochymosin fusion proteins containing the presequence alone, the preproI region, the preproIproII region, and the preproIproII region plus either 14 or 90 amino acid residues of the mature AEP. Each of these fusion proteins was efficiently secreted and released active chymosin on treatment with acid. Thus, the AEP proregion was not necessary for the secretion of prochymosin. In the case of the preproIproII fusion, the AEP secretion and processing signal sequences were shown to be accurately recognized and cleaved. Porcine α-interferon was efficiently secreted when fused to either the preproI or the preproIproII regions of AEP[164,271] although some incorrect processing was reported[133]. The final level of α-interferon secreted was increased about two- to three-fold using an *ARS18* rather than an integrating vector.[271] However, this improvement was less than the six-fold increase found with AEP since α-interferon expression resulted in reduced vector stability. Similar results were obtained for prochymosin which gave only 1·3-fold higher levels using the ARS vector. Regulation of the *XPR2* promoter was also impaired using this vector, since basal AEP levels were found to be 50-fold higher than normal under repressing conditions.

Schizosaccharomyces pombe

Aside from *S.cerevisiae*, the fission yeast, *S.pombe*, is the most intensely studied and well

characterized of the yeast species. This is largely because, like budding yeast, its life cycle and growth characteristics are particularly suited to genetic and biochemical analysis. The early development of a transformation system[22] has led to the cloning and characterization of a large number of genes, and to the development of an array of genetic manipulation techniques comparable to those used in *S. cerevisiae* (for review, see reference 319). Such studies have highlighted how distantly related these two yeasts are. In fact, sequence comparisons show that for many *S.pombe* genes, the mammalian homologues are only marginally more divergent than corresponding *S.cerevisiae* genes,[319] and some are actually less divergent (e.g. *ras*[127]). The utility of *S.pombe* in isolating mammalian genes by complementation of corresponding mutant homologues has been demonstrated,[230] and this remains the most important use of this expression system. In contrast, the use of *S.pombe* for protein production has been limited since little fermentation technology has been developed, and because relatively few inducible promoters were available.

Transformation of *S.pombe* has been described using sphaeroplasts,[22] lithium salts,[165] and electroporation.[179] A highly efficient method uses lipofectin to enhance uptake of DNA by sphaeroplasts.[5] A number of selectable markers have been described (see reference 409) but the *LEU2* and *URA3* genes from *S.cerevisiae*, which complement the *S.pombe leu1* and *ura4* mutations, are most commonly used. The corresponding *S.pombe* genes have also been used,[9,212] and dominant selection using the G418 and bleomycin resistance genes has been demonstrated.[224,323]

S.pombe expression vectors normally contain sequences derived either from the 2μ plasmid of *S. cerevisiae* or from the *S.pombe ars1*. The 2μ sequences that have ARS activity in *S.pombe* do not include the complete 2μ *ORI* and do not depend on any 2μ-encoded function.[132] These vectors have relatively low copy number (5–10 per cell) and mitotic stability (30–45% loss per generation without selection[165]). The *ars1* vectors behave similarly but have higher copy number (about 30/cell[165]). Stability and copy number of *ars1* vectors is greatly enhanced by the presence of an *S.pombe*-derived sequence called *stb*, which appears to provide a partitioning function.[165] These *ars1/stb* vectors, e.g. pFL20,[240] have a copy number of about 80 and are lost at a rate of 13% per generation.[165]

A number of promoters have been used for the expression of foreign genes in *S.pombe*. Promoters from *S.cerevisiae* genes generally function poorly, giving inefficient and aberrant initiation of transcription.[317,318] Nevertheless, the *S.cerevisiae PGK* promoter has been used to express polyoma middle T antigen,[27] and the *ADH1* and *CYC1* promoters used to produce biologically active, secreted antithrombin III.[46] In addition, the *S.cerevisiae* genes for three glycosidases (α-mannosidase, exoglucanase and endochitinase[226]) and the *Saccharomyces diastaticus* glucoamylase gene[112] have all been expressed in *S.pombe* using their respective promoters. However, the SV40 early promoter and the promoter from the *S.pombe* alcohol dehydrogenase (*adh*) gene have been most commonly used. The *adh* gene is constitutive and gives alcohol dehydrogenase levels of about 0·5–2% t.c.p.[320] With the SV40 early promoter transcripts initiate at the same position in *S.pombe* as in mammalian cells[203] although it is weaker than the *adh* promoter. Toyama and Okayama[379] used the chloramphenicol acetyltransferase gene to test a number of other mammalian promoters for activity in *S.pombe*. The human chorionic gonadotrophin α and human cytomegalovirus promoters were about ten-fold stronger than the SV40 promoter and several others, although weaker, were also functional. Other constitutive promoters that have been used are the cauliflower mosaic virus (CaMV) 35S promoter,[145,289] which gave expression levels similar to the *adh* promoter, and a promoter isolated from random *S.pombe* genomic fragments, called 54/1,[224] which gave β-galactosidase levels of 5% t.c.p.

Regulatable expression systems for *S.pombe* have also been described. One uses the promoter from the *S.pombe* fructose bisphosphatase (*fbp*) gene which is expressed at very low levels in cells grown on 8% glucose and is derepressed 100-fold in media containing non-repressing carbon sources.[176] However it should be noted that, even in 8% glucose, the *fbp* promoter is partially induced on entry into stationary phase. A second regulated system uses mammalian glucocorticoid response elements to drive expression. These are induced up to 70-fold by glucocorticoid hormones when the hormone receptor is co-expressed.[287] The fully-induced expression level was about 20% of that using the *adh* promoter. A promising regulated system uses the promoter from a highly-transcribed *nmt1* gene which is strongly repressed by thiamine.[249] This was used to regulate a multi-copy chloramphenicol acetyltransferase (CAT) reporter gene, and gave a 200-fold thiamine-dependent repression of CAT expression.

In addition to the examples already described above, a diverse selection of other heterologous

proteins have been expressed in *S.pombe*. The human liver microsomal enzyme, epoxide hydrolase, was expressed in an active form that could be isolated from *S.pombe* microsomal fractions.[192] Active factor XIIIa was produced at 2 mg/l using a high copy number, *adh* expression vector derived from pFL20.[45] The expression of functional single-chain antibody molecules capable of binding their cognate hapten, the aromatic dye fluorescein, has been described.[88] Jagadish *et al.*[194] expressed the large polyprotein of infectious bursal disease virus, which was processed to give stable VP3 protein. Strasser *et al.*[368] expressed the α-amylase gene from *Schwanniomyces occidentalis*. The bacterial proteins, β-glucuronidase,[289] xylose isomerase[58] and bacterio-opsin[166] have also been produced in *S.pombe*. The latter is correctly processed by cleavage at the N-terminus, is inserted into membranes, and forms photoactive bacteriorhodopsin pigment when cells are grown in the presence of its chromophore, retinal.[166] Another integral membrane protein expressed in active form by *S.pombe* is the glucose transporter from *Arabidopsis thaliana* encoded by the *STP1* gene.[329]

The secretion of foreign proteins is a relatively unexplored area of potential interest since, like higher eukaryotes, *S.pombe* is known to galactosylate glycoproteins (e.g. invertase,[263] acid phosphatase[93]). However, *S.cerevisiae* invertase was not galactosylated when expressed in *S.pombe*,[263] and both this protein and homologous ones are highly glycosylated. In addition, antithrombin III was hyperglycosylated when secreted by *S.pombe*.[46]

PHYSIOLOGY OF FOREIGN GENE EXPRESSION

High-level expression of a foreign gene can place a significant metabolic burden on the host cell, reducing its growth rate and affecting the efficiency of gene expression. Expression of some genes causes a more acute effect, either through a severe effect on metabolism or by direct toxicity. In constitutive systems where growth and expression are linked there is then strong selection for cells with reduced foreign gene expression, so that low-expressing variants arise. Variants may also be selected without growth, for example through effects of the product on cell viability during freezing or long-term storage on agar. Therefore it is highly desirable to use tightly-regulated promoters where the growth and induction phases can be largely separated. This is particularly important with large-scale fermentations where growth is over many generations. With regulated systems only highly toxic products should be problematic, because of their ability to rapidly affect host cell metabolism.

Mechanisms of toxicity

Toxicity is diagnosed by a difficulty in obtaining transformants with constitutive expression vectors or a reduction in growth rate when regulated vectors are induced. Which metabolic process is predominantly affected depends on the foreign gene and vector system. The maintenance of ultra-high-copy 2μ/*LEU2*-d plasmids alone is sufficient to reduce the host cell growth rate,[397,169] whereas no effect would be expected with integrated vectors. Multi-copy promoters may compete for transcription factors and inhibit expression of co-regulated genes: thus multi-copy *GAL7* vectors inhibit *GAL10* gene transcription, and could affect galactose metabolism.[14] High levels of mRNAs containing rare codons could deplete cognate tRNAs and inhibit the translation of host mRNAs containing rare codons. This appears to be the effect of *lacZ* mRNA in yeast, which at high levels is toxic, especially during growth on minimal medium.[292]

Toxicity is a relatively common problem with secreted proteins, especially complex ones. This may be due to the complexity of the secretory pathway, offering a number of possible bottlenecks, or else to the possibility that foreign secreted proteins may be frequently malfolded and cause a blockage in the pathway. Examples of highly toxic secreted proteins are tPA,[169] IGF-1,[343] EBV gp350.[335] In the case of tPA, acid phosphatase secretion is reduced, suggesting a general block in secretion. Membrane proteins, e.g. *E.coli ompA*,[196] influenza virus haemagglutinin,[191] polyoma virus middle T antigen,[27] are particularly likely to be toxic, possibly due to non-specific insertion and disruption of yeast intracellular membranes. This may be the mechanism of toxicity of HBsAg, since constitutively-expressing strains have reduced viability on freezing and thawing (M.A.R., unpublished results).

Certain proteins have specific, acutely toxic effects. As an extreme example, induction of intracellularly expressed A-chain of ricin instantaneously kills yeast cells, and this has been used as a selection for non-toxic A-chain mutants.[125] Over-expression of proteases, such as HIV-1 protease, may be toxic. Mammalian transcriptional *trans*-activators are frequently toxic, presumably through their ability to sequester components of the yeast transcriptional apparatus.

A number of approaches can be used in an attempt to overcome product toxicity. Secretion may be used to segregate the product from the site of toxicity even if it is not normally secreted, e.g. HIV-1 protease.[288] Alternatively, fusion to other proteins may inhibit biological activity and toxicity. The approach of selecting for mutants resistant to toxicity has been used successfully with IGF-1.[347] Toxin-resistant mutants have been used to express diphtheria toxin fragment A, which normally kills yeast cells, and chimaeric diphtheria toxins.[285]

Generation of low-expressing variants

To a certain extent it is possible to limit the generation of variants having reduced expression levels, which can arise with constitutive systems. The most common response to toxicity with 2μ vectors is reduced copy number, which can reach a level as low as 1 per cell even with selection.[35] However, when *LEU2*-d selection is used the drop in copy number that can occur without affecting growth is limited, and other mechanisms such as structural rearrangement predominate. Where the selection marker is flanked by 2μ DNA, recombination can result in its transfer to native 2μ with loss of the expression cassette.[99] One way of preventing this is by using plasmids such as pJDB219 in 2μ-free strains.

Another mechanism for plasmid loss involves homologous recombination between the selection marker and chromosomal allele to generate a prototroph. This can be prevented by using a host strain with a deletion in homologous chromosomal sequences or by use of recombination-deficient mutants (e.g. *rad52*[244]).

An unusual plasmid mutation was observed during the constitutive expression of polyomavirus middle T antigen.[27] Transformants grew very slowly, and there was either plasmid loss or spontaneous generation of plasmids expressing a truncated antigen, which lacked the membrane-spanning domain and was non-toxic. Mutant plasmids arose by deletion of one G.C base pair in a run of nine, causing a frame-shift 5' of the DNA encoding the hydrophobic region.

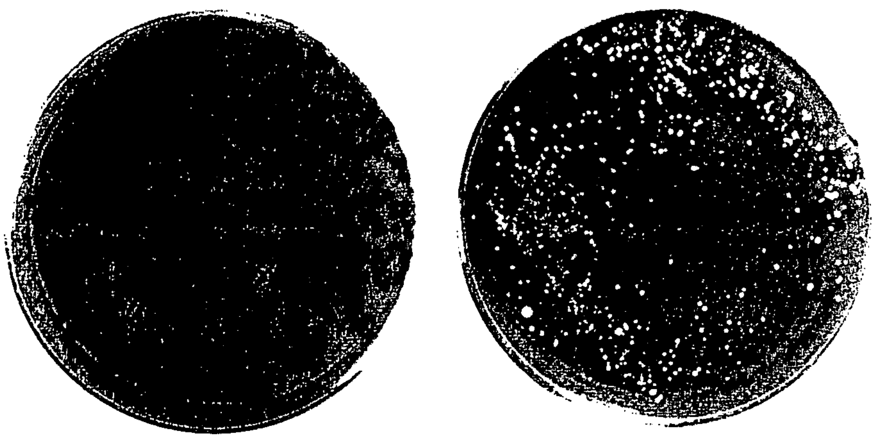
We have observed an example where a mutation in a host cell gene inactivated the promoter used in the expression vector. The system was a *PHO5* vector expressing HBsAg in a constitutive *pho80* background (M.A.R., unpublished results). Mutant non-expressing cells were enriched during freezing and thawing of stock cultures, and out-grew the expressing cells in fermenters, resulting in greatly reduced product yields. Mutant cells were detected using a chromogenic plate assay for secreted acid phosphatase expressed from the chromosomal *PHO5* gene (Figure 12). The problem was eliminated by a combination of measures: (i) using a freezing protocol which gave higher viability to reduce enrichment of mutant cells and (ii) plating out stock cultures and using the plate assay to monitor for non-expressing colonies. The mutants did not arise at detectable frequencies without the enrichment that occurred during low-viability freezing. This example illustrates the type of problem that can occur on scale-up of a constitutive system expressing a mildly toxic product.

Large-scale fermentation and optimization

In the development of a large-scale industrial fermentation process, the first step is transfer to small fermenters. These offer the possibility of controlling a number of culture parameters, resulting in much greater reproducibility and the ability to reach very high cell densities. However, a significant effort is frequently required in order to maximize product yield, though in theory it should be possible to control conditions so as to give higher yields than in shake-flasks. High-density fermentation is generally desirable in order to minimize capital expenditure on fermentation equipment; it is particularly valuable with secretion, where the concentration of secreted protein in the medium can increase almost proportionately to cell density.[213]

In the case of recombinant *S.cerevisiae*, glucose is almost always used as the carbon-source. The two extremes of glucose-utilization are fermentation, which yields ethanol as a by-product, and respiration, which is much more efficient for conversion to biomass. However, growth to high density on glucose is problematic since *S.cerevisiae* has a limited capacity to metabolize glucose oxidatively, accumulating ethanol, which eventually inhibits growth, above a specific growth rate of 0·2 to 0·25 h^{-1}.[117] Secondly, glucose concentrations above approx. 0·1% repress respiration.[298] In aerobic batch cultures, such as shake-flasks, growth on glucose is diauxic: the glucose is rapidly fermented to ethanol and CO_2, and the ethanol is metabolized in a second growth phase once the glucose is exhausted.

In order to prevent the conversion of glucose to ethanol and maximize biomass, bakers' yeast is grown using a fed-batch process: molasses and nutrients are slowly fed to the fermenter, the feed rate increasing as the cell density increases, so as to obtain



A B

Figure 12. Accumulation of non-expressing cells during fermentation of an *S.cerevisiae pho80* strain containing a *PHO5* vector for HBsAg expression. Samples were taken at the end of 10 litre fermenter runs of the constitutively-expressing strain, plated on non-selective agar and assayed for secreted acid phosphatase (Pho phenotype) by overlaying with agar containing a chromogenic substrate. (A) High-yield fermenter run contained Pho⁺ cells only. (B) Low-yield run contained a high proportion of Pho⁻ cells in which the *PHO5* promoter is inactive. Low-yield runs resulted from enrichment of mutant non-expressers during low-viability freezing. Improved freezing protocols resulted in high-yield runs.

an almost constant specific growth rate. Similarly, fed-batch using calculated glucose feed rates has been used for high-density culture of recombinant *S.cerevisiae*.[151,182] More refined methods have also been applied: computer control has been used to couple glucose feed rate to respiratory quotient (RQ = CO_2 evolved/O_2 consumed; maintained at 1·0 to 1·2) using on-line gas analysis,[118,213,384] or to ethanol concentration using a biosensor,[3] resulting in cell densities of 80 to 200 g/l dry mass. These results have been achieved with semi-synthetic media containing protein hydrolysate (casamino acids) or yeast extract.

With an ideal tightly-regulated system, biomass accumulation would be separated from foreign protein induction, and the two could be optimized separately. In constitutive systems this is clearly not possible, but even with regulated systems a combined growth/induction process must usually be developed. We will illustrate some principles of process development by referring to *S.cerevisiae* systems using three types of promoters: constitutive glycolytic, glucose-repressible, and galactose-inducible.

As described above, product toxicity is a serious problem with constitutive systems: Fieschko *et al.*[118] found severe plasmid-loss using a *PGK* vector expressing γ-interferon, and were unable to achieve high cell densities. However, an optimized process has been developed for secretion of pro-urokinase from a super-secreting *pmr1* strain using the *PGK* or *TPI* promoters.[384] Accumulation of secreted product was growth-associated but was not as high in an RQ-controlled fed-batch culture (77 g/l dry mass) as in a perturbed batch fermentation (45 g/l), where glucose was pulsed into a batch culture. This difference may be due to activation of glycolytic promoters by high concentrations of glucose. In batch fermentations the secretion of human lysozyme, using the enolase promoter, was increased four-fold by increasing starting glucose concentration from 2 to 10%, even though growth was unaffected.[185] In view of these observations another approach would be to use glycolytic promoter systems in two-stage fermentations, an initial glucose-limited stage for biomass accumulation followed by an increased glucose feed for induction.

Glucose-repressible expression systems present problems because of the conflicting requirement to maintain high glucose in order to repress the promoter, and to limit glucose in order to maximize biomass. However, a fed-batch process for the production of a SOD-proinsulin fusion protein using a *GAP/ADH2* promoter has been developed, though the cell density achieved was only 30 g/l.[378] Initial experiments indicated that a glucose excess was important to repress the promoter and maintain plasmid stability. Quite high yields of product (>0·5 g/l) were obtained in simple batch cultures with glucose concentration starting at 5%, but biomass was low. A fed-batch procedure was developed using a constant, empirically-determined glucose feed which led to glucose excess during most of the growth phase but to very low glucose at later stages as the cell density increased. Sub-optimal feed-rates would have resulted either in plasmid loss or in incomplete induction. Apparently the residual glucose (20 mg/l) was partially repressing, since changing from glucose to a final ethanol feed yielded a two-fold improvement to 1·2 g/l of product (increasing to 1·6 g/l at 26°C).

The first published study of galactose induction in a high-density fermentation involved the expression of γ-interferon using the *GAP/GAL* hybrid promoter.[118] Although biomass yields of 200 g/l were possible using RQ-controlled glucose-limited fermentations, biomass was restricted to <100 g/l because of O_2-transfer limitation. Three induction methods were compared: (i) 10 g/l pulses of galactose, (ii) replacement of the glucose feed by galactose, (iii) (i) followed by (ii). The third method gave the highest yields and this was rationalized as being due to the partial induction of galactose-metabolizing enzymes during galactose pulsing, preventing temporary carbon-source starvation on switching the feeds. However, in another study galactose was rapidly metabolized in glucose-limited cultures.[3] A further improvement (to 2 g/l of product) was made using a diploid host strain, however the yield of γ-interferon was only 3 to 5% t.c.p. compared to 10% in shake-flasks.

More recently, the production of β-galactosidase was examined using a galactose-inducible, hybrid *CYC1/GAL* promoter.[3] It was found that galactose added to glucose-limited cultures was rapidly metabolized and depleted, since the cells were derepressed, and it was preferable to feed a glucose/galactose mixture so that glucose was preferentially metabolized. In this way β-galactosidase was induced to 7% t.c.p. at a cell density of 100 g/l; addition of a final 50% galactose feed increased the yield to 8%, considerably higher than the level in shake-flasks.

Expression systems based on several yeasts other than *S.cerevisiae* have been scaled-up. A recombinant *K.lactis* strain expressing secreted HSA under the control of the *S.cerevisiae PGK* promoter has been grown to high density (80 to 90 g/l) using glucose-limited fed-batch at the 1000 l scale, yielding HSA at several g/l.[120] A commercial process for the secretion of prochymosin has been scaled-up to 41,000 litres.[391]

P.pastoris is particularly suited to high-density culture and can be grown to >100 g/l on simple defined media containing glycerol without the need for complex fed-batch procedures.[80] The tight regulation of the *AOX1* expression vectors means that the growth phase is identical for different protein products, and scale-up to large fermenters does not affect biomass or product yield.[80] Additionally, contrary to the experience with *S.cerevisiae*, an immediate improvement in percentage yield of foreign protein is generally seen on going from shake-flasks to fermenters.[73,74,306] This appears to be due, at least in part, to the high O_2-demand of the organism, which is not satisfied in shake-flasks.

Figure 13 shows the results of a high-density fermentation using a *Pichia* Mut[s] recombinant containing 14 integrated copies of a tetanus toxin fragment C expression vector.[73] The culture was started with an initial batch growth in glycerol; when this was used up a slow glycerol feed was initiated, during which growth was glycerol-limited and the *AOX1* promoter derepressed, allowing rapid induction during the subsequent methanol feed. Most of the growth occurred during the glycerol feed; product accumulated for about 30 h into the methanol feed, during which time there was no cell doubling, reaching a level of 27% t.c.p. (*cf.* 10% in shake-flasks) or approx. 12 g/l. In contrast, Mut^+ strains continue to grow rapidly during the methanol feed and are fed methanol at a higher rate. Using similar protocols high levels of several commercially-important proteins have been achieved: e.g. SOD (1·3 g/l[375]), HBsAg (0·4 g/l[81]), tumour necrosis factor (10 g/l[360]), *Bordetella pertussis* pertactin (3 g/l[306]), secreted HSA (3 g/l; K. Sreekrishna, personal communication), secreted EGF (0·45 g/l;[74] G. Thill, personal communication). Mut^+ strains can also be grown in continuous culture to increase fermenter productivity.[97]

Apart from the growth and induction protocols, a number of other variables have been found to affect yields and should be looked at systematically. Lower temperatures may increase yield by reducing proteolysis.[378] Added phosphate in the medium has increased heterologous secretion in at least two

a)

b)

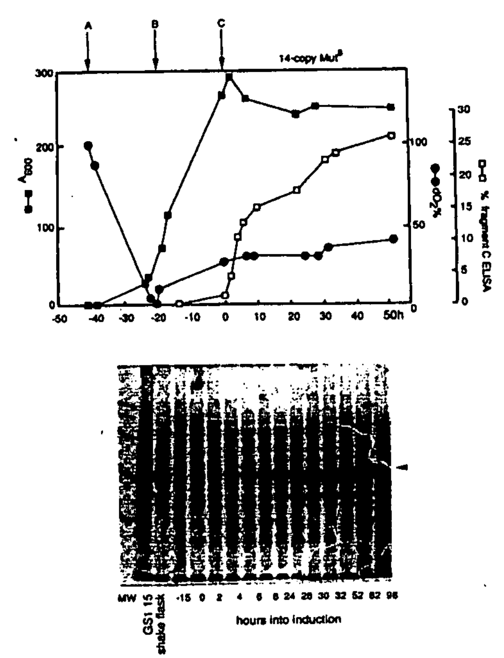


Figure 13. Fermentation of a 14-copy *Pichia* Mut^s transformant expressing tetanus toxin fragment C. (a) Cell density (A_{600}), dissolved oxygen (dO_2), and fragment C as per cent of cell protein are plotted against time (0 h = start of induction). Fermentation conditions were as described elsewhere.[73] Timepoint A, addition of shake-flask inoculum, B, glycerol-limited feed, C, methanol feed. (b) Coomassie blue-stained SDS polyacrylamide gel showing protein extracts from cells taken at different times during induction. The arrow indicates the position of fragment C. For comparison, a sample from the same transformant induced for 48 h in a shake flask is shown in lane 3, and extract from an untransformed control strain is shown in lane 2.

cases.[114,384] Buffering to high pH with *S.cerevisiae*[384] and high or low pH with *P.pastoris*[74] (K. Sreekrishna, personal communication) has reduced proteolysis of secreted proteins. Increased growth rate increases plasmid stability and can result in greater product yield.[82] Finally, different strains of *S.cerevisiae* may have dramatically different growth and induction properties.

CONCLUDING REMARKS

In its early stages the use of yeast for heterologous expression was highly favoured due to the obvious advantages of a microbial eukaryotic system. However, interest waned with the discovery of unforseen problems and with the advent of powerful alternative systems such as baculovirus. Yeast expression

systems are now in a period of resurgence, for a number of reasons. One of the major disadvantages compared to *E.coli* and baculovirus has been the generally lower yield of product, frequently due to the difficulty in obtaining high-level transcription of foreign genes. This problem has now been addressed in *S.cerevisiae* in a number of ways: for example, by over-expressing transcriptional *trans*-activators (e.g. with *ADH2*[291]), by constructing glycolytic promoters with superimposed regulation,[398] and by random screening for super-expressing mutant strains.[354] The use of yeasts such as *Pichia pastoris*, which naturally have powerful, tightly-regulated promoters, has provided an alternative solution. Table 6 (see 'Expression in non-*Saccharomyces* yeasts) shows data for two proteins produced in four expression systems, *S.cerevisiae, P.pastoris, E.coli* and baculovirus, illustrating the comparative strength of the *P.pastoris* multi-copy *AOX1* vector system. Another factor has been the accumulation of experience in the high-density growth and induction of recombinant yeasts, comparing favourably with the difficulties encountered in scaling up other eukaryotic systems. *Pichia* is particularly suited to scale-up and should continue to become more widely used for industrial production.

A number of other areas have shown significant advances, for example the development of more stable 2μ vectors, of rDNA and *Ty* multi-copy integrating vectors, and the use of autoselection markers. The development of episomal vectors for *Kluyveromyces lactis* is also worthy of mention since this yeast appears to be particularly efficient in secretion. However, there are areas where improvements are needed. In *S.cerevisiae* it appears that foreign genes are frequently poorly transcribed using certain promoters, possibly due to a requirement for intragenic yeast sequences (DASs) for maximal transcription. Unfortunately, it is not clear how general the problem is, nor is there yet direct evidence for DASs. In *P.pastoris* there appears to be no such problem using the *AOX1* promoter since very high levels of foreign transcripts can be obtained with multi-copy integrants.[73]

Despite several important successes in secreting proteins, especially unglycosylated polypeptides, this is an area which can present problems. It is still not clear why some foreign proteins are not correctly folded and transported. The current rapid progress in our understanding of protein folding and chaperones may eventually provide explanations and possibly solutions. Recently, however, the most progress in increasing yields of secreted proteins has been made using the powerful empirical approach of random mutagenesis and screening. The secretion of pharmaceutical glycoproteins remains problematic due to the differences in glycosylation between yeast and mammals. Even glycoproteins which are not hyperglycosylated contain antigenic mannose linkages: an engineered strain which could overcome this problem has been reported,[353] though no results on product antigenicity are available yet. An alternative possibility is *P.pastoris* which does not appear to add the antigenic α1,3-linked terminal mannose.[379a] The production of glycoproteins for non-pharmaceutical use, e.g. glucoamylases and xylanases, has been highly successful and presents no such problems.

In the last few years there have been many successes in the production of therapeutic proteins from yeast, for example the recombinant subunit vaccine against hepatitis B virus, human proinsulin, EGF, HSA, etc. There have also been developments in the food industry, such as the experimental use of recombinant yeast secreting glucoamylase in brewing,[157] and the production of chymosin from *K.lactis*.[391] However, in future the requirements of an expression system may be increasingly to provide reagents for research and drug discovery rather than in the production of therapeutic agents. For rapid isolation of such products it may be advantageous to make use of fusion proteins or 'tags' which allow simple affinity purification: a variety of systems are now available.[341,351] Since many pharmacological target proteins may be toxic when expressed to high levels, or may have short half-lives, it may be necessary to accept low yields and concentrate products using an affinity tag.

A highly significant development has come from the realization that many pharmacological target proteins can function in yeast cells *in vivo*. It has been reported that adrenergic receptors can be expressed in yeast and coupled to the signal transduction pathway for α-factor; using a *lacZ* reporter fused to the α-factor regulated *FUS1* promoter, a colorimetric assay that could be used for drug screening was developed.[214] There is intense commercial interest in such systems because of their advantages over mammalian cells in setting up robust drug screens, but there is scope for improvement at present. Other examples of systems that have been or could be used for *in vivo* pharmacological screening are: HIV translational frameshifting,[404] HIV TAT *trans*-activation,[100] steroid receptor binding and *trans*-activation.[293,332]

Finally, yeast expression technology is now important in the genetic analysis of organisms whose genetics are less well developed. A number of mammalian genes have been isolated by their ability to complement mutant yeast homologues, e.g. the human homologue of the *S.pombe cdc2* gene.[230] *S.pombe* may frequently be more suitable than *S.cerevisiae* for screening human cDNA expression libraries because of its closer phylogenetic relationship, and because mammalian expression vectors using the SV40 promoter are active in *S.pombe*.

ACKNOWLEDGEMENTS

We would like to thank the following people who have helped us in writing this review by helpful discussion or by providing us with pre-publication data: A. J. P. Brown, R. G. Buckholz, A. M. Cashmore, I. G. Charles, J. M. Cregg, R. Fleer, M. A. Gleeson, A. R. Goodey, J. Hammond, A. Hinnen, W. E. Lancashire, M. Lockyer, A. J. Makoff, J. Mellor, J. A. H. Murray, I. J. Purvis, J. R. Shuster, K. Sreekrishna, M. J. R. Stark, G. P. Thill, J. F. Tschopp, M. D. Winther.

REFERENCES

1. Adams, B. G. (1972). Induction of galactokinase in *Saccharomyces cerevisiae*: kinetics of induction and glucose effects. *J. Bacteriol.* **111**, 308–315.
2. Adoga, G. and Mattey, M. (1979). Properties of an extracellular peptide with esterase activity produced by *Candida lipolytica*. *FEMS Microbiol. Lett.* **6**, 61–63.
3. Alberghina, L., Porro, D., Martegani, E. and Ranzi, B. M. (1991). Efficient production of recombinant DNA proteins in *Saccharomyces cerevisiae* by controlled high-cell-density fermentation. *Biotechnol. Appl. Biochem.* **14**, 82–92.
4. Allison, D. S. and Young, E. T. (1988). Single-amino-acid substitutions within the signal sequence of yeast prepro-α-factor affect membrane translocation. *Mol. Cell. Biol.* **8**, 1915–1922.
5. Allshire, R. C. (1990). Introduction of large linear minichromosomes into *Schizosaccharomyces pombe* by an improved transformation procedure. *Proc. Natl. Acad. Sci. USA* **87**, 4043–4047.
6. Arffman, A., Aho, S., Torkkeli, H. and Korhola, M. (1990). Isolation and characterization of yeast mutants supersecreting *Trichoderma reesii* endoglucanase I (EGI). *Yeast* **6**, S438.
7. Arima, K., Oshima, T., Kubota, I., Nakamura, N., Mizunaga, T. and Toh-e, A. (1983). The nucleotide sequence of the yeast *PHO5* gene: a putative precursor of repressible acid phosphatase contains a signal peptide. *Nucleic Acids Res.* **11**, 1657–1672.
8. Armstrong, K. A., Som, T., Volkert, F. C., Rose, A. and Broach, J. R. (1989). Propagation and expression of genes in yeast using 2-micron circle vectors. In Barr, P. J., Brake, A. J. and Valenzuela, P. (Eds), *Yeast Genetic Engineering*. Butterworths, pp. 165–192.
9. Bach, M.-L. (1987). Cloning and expression of the OMP decarboxylase gene *ura4* from the yeast *Schizosaccharomyces pombe*. *Curr. Genet.* **12**, 527–534.
10. Baim, S. B., Pietras, D. F., Eustice, D. C. and Sherman, F. (1985). A mutation allowing an mRNA secondary structure diminishes translation of *Saccharomyces cerevisiae* iso-1-cytochrome c. *Mol. Cell. Biol.* **5**, 1839–1846.
11. Baim, S. B. and Sherman, F. (1988). mRNA structures influencing translation in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8**, 1591–1601.
12. Baker, R. T. and Varshavsky, A. (1991). Inhibition of the N-end rule pathway in living cells. *Proc. Natl. Acad. Sci. USA* **88**, 1090–1094.
13. Baker, S. M., Johnston, S. A., Hopper, J. E. and Jaehning, J. A. (1987). Transcription of multiple copies of the yeast *GAL7* gene is limited by specific factors in addition to *GAL4*. *Mol. Gen. Genet.* **208**, 127–134.
14. Baker, S. M., Okkema, P. G. and Jaehning, J. A. (1984). Expression of the *Saccharomyces cerevisiae GAL7* gene on autonomously replicating plasmids. *Mol. Cell. Biol.* **4**, 2062–2071.
15. Baldari, C., Murray, J. A. H., Ghiara, P., Cesareni, G. and Galeotti, C. L. (1987). A novel leader peptide which allows efficient secretion of a fragment of human interleukin 1β in *Saccharomyces cerevisiae*. *EMBO J.* **6**, 229–234.
16. Ballou, C. E. (1970). A study of the immunogenicity of three yeast mannans. *J. Biol. Chem.* **245**, 1197–1203.
17. Barnes, D. A. and Thorner, J. (1986). Genetic manipulation of *Saccharomyces cerevisiae* by use of the *LYS2* gene. *Mol. Cell. Biol.* **6**, 2828–2838.
18. Barr, P. J., Gibson, H. L., Lee-Ng, C. T., Sabin, E. A., Power, M. D., Brake, A. J. and Shuster, J. R. (1987). Heterologous gene expression in the yeast *Saccharomyces cerevisiae*. In Korhola, M. and Nevalainen, H. (Eds), *Industrial Yeast Genetics*, Foundation for Biotechnical and Industrial Fermentation Research, Helsinki, pp. 139–148.
19. Barr, P. J., Steimer, K. S., Sabin, E. A., Parkes, D., George-Nascimento, C., Stephans, J. C., Powers, M. A., Gyenes, A., Van Nest, G. A., Miller, E. T., Higgins, K. W. and Luciw, P. A. (1987). Antigenicity and immunogenicity of domains of the human immunodeficiency virus (HIV) envelope protein expressed in the yeast *Saccharomyces cerevisiae*. *Vaccine* **5**, 90–101.

20. Bartel, B., Wunning, I. and Varshavsky, A. (1990). The recognition component of the N-end rule pathway. *EMBO J.* **9**, 3179–3189.
21. Bathurst, I. C., Brennan, S. O., Carrell, R. W., Cousens, L. S., Brake, A. J. and Barr, P. J. (1987). Yeast KEX2 protease has the properties of a human proalbumin converting enzyme. *Science* **235**, 348–350.
22. Beach, D. and Nurse, P. (1981). High frequency transformation of the fission yeast *Schizosaccharomyces pombe*. *Nature* **290**, 140–142.
23. Beesley, K. M., Francis, M. J., Clarke, B. E., Beesley, J. E., Dopping-Hepenstal, P. J. C., Clare, J. J., Brown, F. and Romanos, M. A. (1990). Expression in yeast of amino-terminal peptide fusions to hepatitis B core antigen and their immunological properties. *Bio/Technology* **8**, 644–649.
24. Beggs, J. D. (1978). Transformation of yeast by a replicating hybrid plasmid. *Nature* **275**, 104–109.
25. Beggs, J. D., van den Berg, J., Van Ooyen, A. and Weissman, C. (1980). Abnormal expression of chromosomal rabbit β-globin gene in *Saccharomyces cerevisiae*. *Nature* **283**, 835–840.
26. Beier, D. R., Sledziewski, A. and Young, E. T. (1985). Deletion analysis identifies a region, upstream of the *ADH2* gene of *Saccharomyces cerevisiae*, which is required for *ADR1*-mediated derepression. *Mol. Cell. Biol.* **5**, 1743–1749.
27. Belsham, G. J., Barker, D. G. and Smith, A. E. (1986). Expression of polyoma virus middle-T antigen in *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **156**, 413–421.
28. Bennetzen, J. L. and Hall, B. D. (1982). Codon selection in yeast. *J. Biol. Chem.* **257**, 3026–3031.
29. Benton, B. M., Eng, W.-K., Dunn, J. J., Studier, F. W., Sternglanz, R. and Fisher, P. A. (1990). Signal-mediated import of bacteriophage T7 RNA polymerase into the *Saccharomyces cerevisiae* nucleus and specific transcription of target genes. *Mol. Cell. Bio.* **10**, 353–360.
30. Bianchi, M. M., Falcone, C., Jie, C. X., Wesolowski-Louvel, M., Frontali, L. and Fukuhara, H. (1987). Transformation of the yeast *Kluyveromyces lactis* by new vectors derived from the 1·6μm circular plasmid pKD1. *Curr. Genet.* **12**, 185–192.
31. Biemans, R., Thines, D., Rutgers, T., De Wilde, M. and Cabezón, T. (1991). The large surface protein of hepatitis B virus is retained in the yeast endoplasmic reticulum and provokes its unique enlargement. *DNA Cell Biol.* **10**, 191–200.
32. Bijvoet, J. F. M., van der Zanden, A. L., Goosen, N., Brouwer, J. and van de Putte, P. (1991). DNA insertions in the silent regions of the 2μm plasmid of *Saccharomyces cerevisiae* influence plasmid stability. *Yeast* **7**, 347–356.
33. Bitter, G. A., Chang, K. K. H. and Egan, K. M. (1991). A multi-component upstream activation sequence of the *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase gene promoter. *Mol. Gen. Genet.* **231**, 22–32.
34. Bitter, G. A., Chen, K. K., Banks, A. R. and Lai, P.-H. (1984). Secretion of foreign proteins from *Saccharomyces cerevisiae* directed by α-factor gene fusions. *Proc. Natl. Acad. Sci. USA* **81**, 5330–5334.
35. Bitter, G. A. and Egan, K. M. (1988). Expression of interferon-gamma from hybrid yeast *GPD* promoters containing upstream regulatory sequences from the *GAL1-GAL10* intergenic region. *Gene* **69**, 193–207.
36. Blachly-Dyson, E. and Stevens, T. H. (1987). Yeast carboxypeptidase Y can be translocated and glycosylated without its amino-terminal signal sequence. *J. Cell Biol.* **104**, 1183–1191.
37. Boeke, J. D., LaCroute, F. and Fink, G. R. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* **197**, 345–346.
38. Boeke, J. D., Xu, H. and Fink, G. R. (1988). A general method for the chromosomal amplification of genes in yeast. *Science* **239**, 280–282.
39. Bowdish, K., Tang, Y., Hicks, J. B. and Hilvert, D. (1991). Yeast expression of a catalytic antibody with chorismate mutase activity. *J. Biol. Chem.* **266**, 11901–11908.
40. Braddock, M., Chambers, A., Wilson, W., Esnouf, M. P., Adams, S. E., Kingsman, A. J. and Kingsman, S. M. (1989). HIV-1 TAT 'activates' presynthesised RNA in the nucleus. *Cell* **58**, 269–279.
41. Brake, A. J. (1989). Secretion of heterologous proteins directed by the yeast α-factor leader. In Barr, P. J., Brake, A. J. and Valenzuela, P. (Eds.), *Yeast Genetic Engineering*, Butterworths, pp. 269–280.
42. Brake, A. J., Merryweather, J. P., Coit, D. G., Heberlein, U. A., Masiarz, F. R., Mullenbach, G. T., Urdea, M. S., Valenzuela, P. and Barr, P. J. (1984). α-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **81**, 4642–4646.
43. Brierley, R. A., Siegel, R. S., Bussineau, C. M., Craig, W. S., Holtz, G. C., Davis, G. R., Buckholtz, R. G., Thill, G. P., Wondrack, L. M., Digan, M. E., Harpold, M. M., Lair, S. V., Ellis, S. B. and Williams, M. E. (1990). Mixed feed recombinant yeast fermentation. U.S. patent WO 9003431.
44. Broach, J. R. (1983). Construction of high copy number yeast vectors using 2μm circle sequences. *Meth. Enzymol.* **101**, 307–325.
45. Broker, M. and Bauml, O. (1989). New expression vectors for the fission yeast *Schizosaccharomyces pombe*. *FEBS Letts* **248**, 105–110.
46. Broker, M., Ragg, H. and Karges, H. E. (1987). Expression of human antithrombin III in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Biochim. Biophys. Acta.* **908**, 203–213.

47. Brown, A. J. P. (1989). Messenger RNA stability in yeast. *Yeast* **5**, 239–257.
48. Brown, A. J. P., Purvis, I. J., Santiago, T. C., Bettany, A. J. E., Loughlin, L. and Moore, J. (1988). Messenger RNA degradation in *Saccharomyces cerevisiae*. *Gene* **72**, 151–160.
49. Bussey, H. (1988). Proteases and the processing of precursors to secreted proteins in yeast. *Yeast* **4**, 17–26.
50. Bussey, H. and Meaden, P. (1985). Selection and stability of yeast transformants expressing cDNA of an M1 killer toxin-immunity gene. *Curr. Genet.* **9**, 285–291.
51. Bussey, H., Steinmetz, O. and Saville, D. (1983). Protein secretion in yeast: Two chromosomal mutants that oversecrete killer toxin in *Saccharomyces cerevisiae*. *Curr. Genet.* **7**, 449–456.
52. Butler, J. S., Sadhale, P. P. and Platt, T. (1990). RNA processing in vitro produces mature 3' ends of a variety of *Saccharomyces cerevisiae* mRNAs. *Mol. Cell. Biol.* **10**, 2599–2605.
53. Cabezón, T, De Wilde, M., Herion, P., Loriau, R. and Bollen, A. (1984). Expression of human α1-antitrypsin cDNA in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **81**, 6594–6598.
54. Campbell, J. L. (1983). Yeast DNA replication. In Setlow, J. K. and Hollaender, A. (Eds), *Genetic Engineering Principles and Methods, Vol. 5*. Plenum, New York, pp. 109–146.
55. Caplan, S., Green, R., Rocco, J. and Kurjan, J. (1991). Glycosylation and structure of the yeast *MFα1* α-factor precursor is important for efficient transport through the secretory pathway. *J. Bact.* **173**, 627–635.
56. Carlson, J. R. (1988). A new means of inducibly inactivating a cellular protein. *Mol. Cell. Biol.* **8**, 2638–2646.
57. Cashmore, A. M., Albury, M. S., Hadfield, C. and Meacock, P. M. (1986). Genetic analysis of partitioning functions encoded by the 2μm circle of *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **203**, 154–162.
58. Chan, E. C., Ueng, P. P. and Chen, L. (1986). D-Xylose fermentation to ethanol by *Schizosaccharomyces pombe* cloned with xylose isomerase gene. *Biotechnol. Letts.* **8**, 231–234.
59. Chang, C. N., Matteucci, M., Perry, L. J., Wulf, J. J., Chen, C. Y. and Hitzeman, R. A. (1986). *Saccharomyces cerevisiae* secretes and correctly processes human interferon hybrid proteins containing yeast invertase signal peptides. *Mol. Cell. Biol.* **6**, 1812–1819.
60. Chang, Y.-H., Teichert, U. and Smith, J. A. (1990). Purification and characterisation of a methionine aminopeptidase from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**, 19892–19897.
60a. Charles, I. G., Rodgers, B. C., Makoff, A. J., Chatfield, S. N., Slater, D. E. and Fairweather, N. F. (1991). Synthesis of tetanus toxin fragment C in insect cells by use of a baculovirus expression system. *Infect. Immun.* **59**, 1627–1632.
61. Chattoo, B. B., Sherman, F., Azubalis, D. A., Fjellstadt, T. A., Mehvert, D. and Oghur, M. (1979). Selection of *lys2* mutants of the yeast *Saccharomyces cerevisiae* by the utilisation of α-aminoadipate. *Genetics* **93**, 51–65.
62. Chen, C. Y., Oppermann, H. and Hitzeman, R. A. (1984). Homologous versus heterologous gene expression in the yeast, *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **12**, 8951–8970.
63. Chen, W., Tabor, S. and Struhl, K. (1987). Distinguishing between mechanisms of eukaryotic transcriptional activation with bacteriophage T7 RNA polymerase. *Cell* **50**, 1047–1055.
64. Chen, X. J., Bianchi, M. M., Suda, K. and Fukuhara, H. (1989). The host range of the pKD1-derived plasmids in yeast. *Curr. Genet.* **16**, 95–98.
65. Chen, X. J., Saliola, M., Falcone, C., Bianchi, M. M. and Fukuhara, H. (1986). Sequence organisation of the circular plasmid pKD1 from the yeast *Kluyveromyces drosophilarum*. *Nucleic Acids Res.* **14**, 4471–4481.
66. Cheng, S. C. and Ogrydziak, D. M. (1986). Extracellular RNase produced by *Yarrowia lipolytica*. *J. Bacteriol.* **168**, 1433–1440.
67. Chinery, S. A. and Hinchcliffe, E. (1989). A novel class of vector for yeast transformation. *Curr. Genet.* **16**, 21–25.
68. Chisholm, V., Chen, C. Y., Simpson, N. J. and Hitzeman, R. A. (1990). Molecular and genetic approach to enhancing protein secretion. *Meth. Enzymol.* **185**, 471–482.
69. Chlebowicz-Sledziewska, E. and Sledziewski, A. Z. (1985). Construction of multicopy yeast plasmids with regulated centromere function. *Gene* **39**, 25–31.
70. Cigan, A. M. and Donahue, T. F. (1987). Sequence and structural features associated with translational initiator regions in yeast—a review. *Gene* **59**, 1–18.
71. Cigan, A. M., Pabich, E. K. and Donahue, T. F. (1988). Mutational analysis of the *HIS4* translational initiator region in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8**, 2964–2975.
72. Clancy, S., Mann, C., Davis, R. W. and Calos, M. P. (1984). Deletion of plasmid sequences during *Saccharomyces cerevisiae* transformation. *J. Bacteriol.* **159**, 1065–1067.
73. Clare, J. J., Rayment, F. B., Ballantine, S. P., Sreekrishna, K. and Romanos, M. A. (1991). High-level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. *Bio/Technology* **9**, 455–460.
74. Clare, J. J., Romanos, M. A., Rayment, F. B., Rowedder, J. E., Smith, M. A., Payne, M. M., Sreekrishna, K. and Henwood, C. A. (1991). Production of mouse epidermal growth factor in yeast: high-level secretion using *Pichia pastoris* strains containing multiple gene copies. *Gene* **105**, 205–212.

75. Clarke, L. and Carbon, J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* **257**, 504–509.
76. Couderc, R. and Baratti, J. (1980). Oxidation of methanol by the yeast *Pichia pastoris*. Purification and properties of alcohol oxidase. *Agric. Biol. Chem.* **44**, 2279–2289.
77. Cousens, D. J., Wilson, M. J. and Hinchcliffe, E. (1990). Construction of a regulated PGK expression vector. *Nucleic Acids Res.* **18**, 1308.
78. Cousens, L. S., Shuster, J. R., Gallegos, C., Ku, L., Stempien, M. M., Urdea, M. S., Sanchez-Pescador, R., Taylor, A. and Tekamp-Olson, P. (1987). High level expression of proinsulin in the yeast, *Saccharomyces cerevisiae*. *Gene* **61**, 265–275.
79. Cregg, J. M., Barringer, K. J., Hessler, A. Y. and Madden, K. R. (1985). *Pichia pastoris* as a host system for transformations. *Mol. Cell Biol.* **5**, 3376–3385.
80. Cregg, J. M. and Madden, K. N. (1987). Development of transformation systems and construction of methanol-utilisation-defective mutants of *Pichia pastoris* by gene disruption. In Stewart, G. G, Russell, I., Klein, R. D. and Hiebsch, R. R. (Eds), *Biological Research on Industrial Yeast*, Vol II. CRC Press, pp. 1–18.
81. Cregg, J. M., Tschopp, J. F., Stillman, C., Siegel, R., Akong, M., Craig, W. S., Buckholz, R. G., Madden, K. R., Kellaris, P. A., Davies, G. R., Smiley, B. L., Cruze, J., Torregrossa, R., Velicelebi, G. and Thill, G. P. (1987). High-level expression and efficient assembly of hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris*. *Bio/Technology* **5**, 479–485.
82. Da Silva, N. A. and Bailey, J. E. (1991). Influence of dilution rate and induction of cloned gene expression in continuous fermentations of recombinant yeast. *Biotechnol. Bioeng.* **37**, 309–317.
83. Danhash, N., Gardner, D. C. J. and Oliver, S. G. (1991). Heritable damage to yeast caused by transformation. *Bio/Technology* **9**, 179–182.
84. Das, S. and Hollenberg, C. P. (1982). A high-frequency transformation system based on the yeast *Kluyveromyces lactis*. *Curr. Genet.* **6**, 123–128.
85. Davidow, L. S., Apostolakis, D., O'Donnell, M. M., Procter, A. R., Ogrydziak, D. M., Wing, R. A., Stasko, I. and DeZeeuw, J. R. (1985). Integrative transformation of the yeast *Yarrowia lipolytica*. *Curr. Genet.* **10**, 39–48.
86. Davidow, L. S., Franke, A. E., DeZeeuw, J. R. (1987). Eur. Patent Application 220864.
87. Davidow, L. S., O'Donnell, M. M., Kaczmarek, F. S., Pereira, D. A., DeZeeuw, J. R. and Franke, A. E. (1987). Cloning and sequencing of the alkaline extracellular protease gene of *Yarrowia lipolytica*. *J. Bacteriol.* **169**, 4621–4629.
88. Davis, G. T., Bedzyk, W. D. Voss, E. W. and Jacobs, T. W. (1991). Single chain antibody (SCA) encoding genes: one step construction and expression in eukaryotic cells. *Bio/Technology* **9**, 165–169.
89. De Baetselier, A., Vasavada, A., Dohet, P., Ha-Thi, V., De Beukelaer, M., Erpicum, T., De Clerk, L., Hanotier, J. and Rosenberg, S. (1991). Fermentation of a yeast producing *A.niger* glucose oxidase: Scale-up, purification and characterisation of the recombinant enzyme. *Bio/Technology* **9**, 559–561.
90. De Louvencourt, L., Fukuhara, H., Heslot, H. and Wesolowski, M. (1983). Transformation of *Kluyveromyces lactis* by killer plasmid DNA. *J. Bacteriol.* **154**, 737–742.
90a. Demolder, J., Fiers, W. and Contreras, R. (1992). Efficient synthesis of secreted murine interleukin-2 by *Saccharomyces cerevisiae*: influence of 3'-untranslated regions and codon usage. *Gene* **111**, 207–213.
91. De Nobel, J. G. and Barnett, J. A. (1991). Passage of molecules through yeast cell walls: a brief essay-review. *Yeast* **7**, 313–323.
92. DeZeeuw, J. and Stasko, I. (1983). U.S. Patent 4407957.
93. DeZeeuw, J. and Tynan, E. J. (1973). U.S. Patent 3736229.
94. DeZeeuw, J. and Tynan, E. J. (1973). U.S. Patent 3736917.
95. Dibenedetto, G. and Cozzani, I. (1975). Non-specific acid phosphatase from *Schizosaccharomyces pombe*. Purification and physical chemical properties. *Biochemistry* **14**, 2847–2852.
96. Dice, F. (1987). Molecular determinants of protein half-lives in eukaryotic cells. *FASEB J.* **1**, 349–357.
96a. Dickson, R. C. and Riley, M. I. (1989) The lactose-galactose regulon of *Kluyveromyces lactis*. In Barr, P. J., Brake, A. J. and Valenzuela, P. (Eds), *Yeast Genetic Engineering*, Butterworths, pp. 19–40.
97. Digan, M. E., Lair, S. V., Brierly, R. A., Siegel, R. S., Williams, M. E., Ellis, S. B., Kellaris, P. A., Provow, S. A., Craig, W. S., Velicelebi, G., Harpold, M. M. and Thill, G. P. (1989). Continuous production of a novel lysosyme via secretion from the yeast, *Pichia pastoris*. *Bio/Technology* **7**, 160–164.
98. Dobson, M. J., Futcher, A. B. and Cox, B. S. (1980). Control of recombination within and between DNA plasmids of *Saccharomyces cerevisiae*. *Curr. Genet.* **2**, 193–200.
99. Dobson, M. J., Futcher, A. B. and Cox, B. S. (1980). Loss of 2μm DNA from *Saccharomyces cerevisiae* transformed with the chimaeric plasmid pJDB219. *Curr. Genet.* **2**, 201–205.
100. Dykes, C. W., Ernst, J. F. and Hobden, A. N. (1991). Eur. Patent Application 91300903.1
101. Ecker, D. J., Stadel, J. M., Butt, T. R., Marsh, J. A., Monia, B. P., Powers, D. A., Gorman, J. A., Clark, P. E., Warren, F., Shatzman, A. and Crooke, S. T. (1989). Increasing gene expression in yeast by fusion to ubiquitin. *J. Biol. Chem.* **264**, 7715–7719.

102. Egel-Mitani, M., Flygenring, H. P. and Hansen, M. T. (1990). A novel aspartyl protease allowing *KEX2*-independent *MFα* propheromone processing in yeast. *Yeast* **6**, 127–137.
103. Eggeling, L. and Samm, H. (1978). Direct enzymatic assay for alcohol oxidase, alcohol dehydrogenase and formaldehyde dehydrogenase in colonies of *Hansenula polymorpha*. *Appl. Environ. Microbiol.* **42**, 268–269.
104. Egli, T., Van Dijken, J. P., Veenhuis, M., Harder, W. and Feichter, A. (1980). Methanol metabolism in yeasts: Regulation of the synthesis of catabolite enzymes. *Arch. Microbiol.* **124**, 115–121.
105. Eisen, A., Taylor, W. E., Blumberg, H. and Young, E. T. (1988). The yeast regulatory protein ADR1 binds in a zinc-dependent manner to the upstream activating sequence of *ADH2*. *Mol. Cell. Biol.* **8**, 4552–4556.
106. Elliott, S., Giffin, J., Suggs, S., Lau, E. P. and Banks, A. R. (1989). Secretion of glycosylated human erythropoetin from yeast directed by the α-factor leader region. *Gene* **79**, 167–180.
107. Ellis, S. B., Brust, P. F., Koutz, P. J., Waters, A. F., Harpold, M. M. and Gingeras, T. R. (1985). Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. *Mol. Cell. Biol.* **5**, 1111–1121.
108. Elroy-Stein, O., Fuerst, T. R. and Moss, B. (1989). Cap-independent translation of mRNA conferred by encephalomyocarditis virus 5' sequence improves the performance of the vaccinia virus/bacteriophage T7 hybrid expression system. *Proc. Natl. Acad. Sci. USA* **86**, 6126–6130.
109. Erhart, E. and Hollenberg, C. P. (1983). The presence of a defective *LEU2* gene in 2μm DNA recombinant plasmids of *Saccharomyces cerevisiae* is responsible for curing and high copy number. *J. Bacteriol.* **156**, 625–635.
110. Ernst, J. F. (1986). Improved secretion of heterologous proteins by *Saccharomyces cerevisiae*: effects of promoter substitution in alpha-factor fusions. *DNA* **5**, 483–491.
111. Ernst, J. F. (1988). Efficient secretion and processing of heterologous proteins in *Saccharomyces cerevisiae* is mediated solely by the pre-segment of α-factor precursor. *DNA* **7**, 355–360.
112. Erratt, J. A. and Nasim, A. (1986). Cloning and expression of a *Saccharomyces diastaticus* glucoamylase gene in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *J. Bacteriol.* **166**, 484–490.
113. Etcheverry, T. (1990). Induced expression using yeast copper metallothionein promoter. *Meth. Enzymol.* **185**, 319–329.
114. Faria, J. B., Castilho-Valavicius, B. and Schenberg, A. C. (1989). Characterisation of secretion of mouse pancreatic α-amylase by *S. cerevisiae*. Abstract 61B, Yeast Genetics and Molecular Biology Meeting, Genetics Society of America, Atlanta, GA.
115. Fellinger, A. J., Verkabel, J. A., Veale, R. A., Sudbery, P. E., Bom, I. J., Overbeeke, N. and Verrips, C. T. (1991). Expression of the α-galactosidase from *Cyamopsis tetragonoloba* (guar) by *Hansenula polymorpha*. *Yeast* **7**, 463–473.
116. Ferro-Novick, S., Hansen, W., Schauer, I. and Schekman, R. (1984). Genes required for completion of import of proteins into the endoplasmic reticulum in yeast. *J. Cell Biol.* **98**, 44–53.
117. Fiechter, A., Fuhrmann, G. F. and Kappeli, C. (1981). Regulation of glucose metabolism in growing yeast cells. *Adv. Microbial Physiol.* **22**, 123–183.
118. Fieschko, J. C., Egan, K. M., Ritch, T., Koski, R. A., Jones, M. and Bitter, G. A. (1987). Controlled expression and purification of human immune interferon from high-cell-density fermentations of *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **29**, 113–121.
119. Fleer, R., Chen, X. J., Amellal, N., Yeh, Y., Fournier, A., Guinet, F., Gault, N., Faucher, D., Folliard, F., Fukuhara, H. and Mayaux, J-F. (1991). High-level secretion of correctly processed recombinant human interleukin-1β in *Kluyveromyces lactis*. *Gene*, in press.
120. Fleer, R., Yeh, P., Amellal, N., Maury, I., Fournier, A., Bacchetta, F., Baduel, P., Jung, G., L'Hote, H., Becquart, J., Fukuhara, H. and Mayaux, J. F. (1991). Stable multicopy vectors for high-level secretion of recombinant human serum albumin by *Kluyveromyces* yeasts. *Bio/Technology* **9**, 968–975.
121. Fleig, U. N., Pridmore, R. D. and Philippsen, P. (1986). Construction of *LYS2* cartridges for use in genetic manipulation of *Saccharomyces cerevisiae*. *Gene* **46**, 237–245.
122. Fogel, S. and Welch, J. W. (1982). Tandem gene amplification mediates copper resistance in yeast. *Proc. Natl. Acad. Sci. USA* **79**, 5342–5346.
123. Fournier, P., Guyaneux, L., Chasles, M. and Gaillardin, C. (1991). Scarcity of ARS sequences isolated in a morphogenesis mutant of the yeast *Yarrowia lipolytica*. *Yeast* **7**, 25–36.
124. Franke, A. E., Kaczmark, F. S., Eisenhard, M. E., Geoghehan, K. F., DeZeeuw, J. R., O'Donnell, M. M., Gollaher, M. G. and Davidow, L. S. (1988). Expression and secretion of bovine prochymosin in *Yarrowia lipolytica*. *Develop. Indust. Microbiol.* **29**, 43–57.
125. Frankel, A., Schlossman, D., Welsh, P., Hertler, A., Withers, D. and Johnston, S. (1989). Selection and characterisation of ricin toxin A-chain mutations in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **9**, 415–420.
126. Freedman, R. B. (1989). Protein disulphide isomerase: multiple roles in the modification of nascent secretory proteins. *Cell* **57**, 1069–1072.
127. Fukui, Y. and Kaziro, Y. (1985). Molecular cloning and sequence analysis of a *ras* gene from *Schizosaccharomyces pombe*. *EMBO J.* **4**, 687–691.

128. Futcher, A.B. (1988). The 2μm circle plasmid of *Saccharomyces cerevisiae*. *Yeast* **4**, 27–40.
129. Futcher, A. B. and Cox, B. S. (1983). Maintenance of the 2μm circle plasmid in populations of *Saccharomyces cerevisiae*. *J. Bacteriol.* **154**, 612–622.
130. Futcher, A. B. and Cox, B. S. (1984). Copy number and stability of 2-μm circle-based artificial plasmids of *Saccharomyces cerevisiae*. *J. Bacteriol.* **157**, 283–290.
131. Gabrielsen, O. S., Reppe, S., Sæther, O., Blingsmo, O. R., Sletten, K., Gordeladze, J. O., Høgset, A., Gautvik, V. T., Alestrøm, P., Øyen, T. B. and Gautvik, K. M. (1990). Efficient secretion of human parathyroid hormone by *Saccharomyces cerevisiae*. *Gene* **90**, 255–262.
132. Gaillardin, C., Fournier, P., Budar, F., Kudla, B., Gerbaud, C. and Heslot, H. (1983). Replication and recombination of 2μ DNA in *Schizosaccharomyces pombe*. *Curr. Genet.* **7**, 245–253.
133. Gaillardin, C., Fournier, P., Nicaud, J. M., Fabre, E., Lopez, M. C. and Tharaud, C. (1990). Genetic engineering in *Yarrowia lipolytica*. In Heslot, H., Davies, J., Florent, J., Bobichon, L., Durand, G. and Penasse, L. (Eds). *Proceedings of the 6th International Symposium on Genetics of Microorganisms*, Vol II. Societe Francaise de Microbiologie, Paris, France, pp. 509–518.
134. Gaillardin, C. and Ribet, A-M. (1987). *LEU2*-directed expression of β-galactosidase activity and phleomycin resistance in *Yarrowia lipolytica*. *Curr. Genet.* **11**, 369–375.
135. Gaillardin, C., Ribet, A. M. and Heslot, H. (1985). Integrative transformation of the yeast *Yarrowia lipolytica*. *Curr. Genet.* **10**, 49–58.
136. Gallili, G. and Lampen, J. O. (1977). Large and small invertase and the yeast cell cycle: pattern of synthesis and sensitivity to tunicamycin. *Biochem. Biophys. Acta* **475**, 113–122.
137. Gellison, G., Janowicz, Z. J., Merckelbach, A., Piontek, M., Keup, P., Weydemann, U., Hollenberg, C. P. and Strasser, W. M. (1991). Heterologous gene expression in *Hansenula polymorpha*: Efficient secretion of glucoamylase. *Bio/Technology* **9**, 291–295.
138. George-Nascimento, C., Gyenes, A., Halloran, S. M., Merryweather, J. P., Steimer, K. S., Masiarz, F. R. and Randolph, A. (1988). Characterisation of recombinant human epidermal growth factor produced in yeast. *Biochemistry* **27**, 797–802.
139. Gerbaud, C., Fournier, P., Blanc, H., Aigle, M., Heslot, H. and Guerineau, M. (1979). High frequency of yeast transformation by plasmids carrying part or entire 2-μm yeast plasmid. *Gene* **5**, 233–253.
140. Gething, M.-J. and Sambrook, J. (1992). Protein folding in the cell. *Nature* **355**, 33–45.
141. Gill, G. S., Zaworski, P. G., Marotti, K. R. and Rehberg, E. F. (1990). A novel screening system for yeast strains capable of secreting tissue plasmino gen activator. *Bio/Technology* **8**, 956–958.
142. Giuseppin, M. L. F., van Eijk, H. M. J. and Bes B. C. M. (1988). Molecular regulation of methano oxidase activity in continuous cultures of *Hansenul. polymorpha*. *Biotechnol. Bioeng.* **32**, 577–583.
143. Gleeson, M. A, Ortori, G. S. and Sudbery, P. E (1986). Transformation of the methylotrophic yeas *Hansenula polymorpha*. *J. Gen. Microbiol.* **132** 3459–3465.
144. Gleeson, M. A. and Sudbery, P. E. (1989). Th methylotrophic yeasts. *Yeast* **4**, 1–15.
145. Gmunder, H., and Kohli, J. (1989). Cauliflowe mozaic virus promoters direct the efficient e) pression of a bacterial G418 resistance gene i *Schizosaccharomyces pombe*. *Molec. Gen. Gene* **220**, 95–101.
146. Goff, C. G., Moir, D. T., Kohno, T., Graviu T. C., Smith, R. A., Yamasaki, E. and Tauntor Rigby, A. (1984). The expression of calf proch) mosin in *Saccharomyces cerevisiae*. *Gene* **27**, 35–4(
147. Goodey, A. R., Doel, S. M., Piggott, J. R., Watsoı M. E. E., Zealey, G. R., Cafferkey, R. and Carte B. L. A. (1986). The selection of promoters fc the expression of heterologous genes in the yea: *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **20** 505–511.
148. Green, R., Schaber, M. D., Shields, D. an Kramer, R. (1986). Secretion of somatostatin b *Saccharomyces cerevisiae*. Correct processing of a α-factor-somatostatin hybrid. *J. Biol. Chem.* **26** 7558–7565.
149. Grinna, L. S. and Tschopp, J. F. (1989). Size distr bution and general structural features of N-linke oligosaccharides from the methylotrophic yeas *Pichia pastoris*. *Yeast* **5**, 107–115.
150. Gritz, L. and Davies, J. (1983). Plasmid encode hygromycin B resistance: the sequence of hygr mycin B phosphotransferase and its expression i *Escherichia coli* and *Saccharomyces cerevisia* *Gene* **25**, 179–188.
151. Gu, M. B., Park, M. H. and Kim, D.-I. (1991 Growth control in fed-batch cultures of recombinaı *Saccharomyces cerevisiae* producing hepatitis B su face antigen (HBsAg). *Appl. Microbiol. Biotechn* **35**, 46–50.
152. Guisez, Y., Tison, B., Vandekerckhove, . Demolder, J., Bauw, G., Haegeman, G., Fiers, V and Contreras, R. (1991). Production and puri cation of recombinant human interleukin-6 secret by the yeast *Saccharomyces cerevisiae*. *Eur. Biochem.* **198**, 217–222.
153. Hadfield, C., Cashmore, A. M. and Meacock, P. (1986). An efficient chloramphenicol-resistan marker for *Saccharomyces cerevisiae* and *Escherich coli*. *Gene* **45**, 149–158.
154. Hadfield, C., Jordan, B. E., Mount, R. C., Pretoriı G. H. J. and Burak, E. (1990). G418-resistance as

cloning marker and reporter for gene expression in *Saccharomyces cerevisiae*. *Curr. Genet.* **18**, 303–313.

154a. Hagenson, M. J., Holden, K. A., Parker, K. A., Wood P. J., Cruze, J. A., Fuke, M., Hopkins, T. R. and Stroman, D. W. (1989). Expression of streptokinase in *Pichia pastoris* yeast. *Enzyme Microb. Technol.* **11**, 650–656.

155. Haguenauer-Tsapis, R. and Hinnen, A. (1984). A deletion that includes the signal peptidase cleavage site impairs processing, glycosylation and secretion of cell surface yeast acid phosphatase. *Mol. Cell. Biol.* **4**, 2668–2675.

156. Hallewell, R. A., Mills, R., Tekamp-Olson, P., Blacher, R., Rosenberg, S., Otting, F., Masiarz, F. R. and Scandella, C. J. (1987). Amino terminal acetylation of authentic human Cu, Zn superoxide dismutase produced in yeast. *Bio/Technology* **5**, 363–366.

157. Hammond, J. (1991). Developments in breeding new brewer's yeasts. *Brewing and Distilling Int.*, April, pp. 16–17.

158. Hashimoto, H., Kikuchi, Y., Nogi, Y. and Fukasawa, T. (1983). Regulation of expression of the galactose gene cluster in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **191**, 31–38.

159. Henderson, R. C. A., Cox, B. S. and Tubb, R. (1985). The transformation of brewing yeasts with a plasmid containing the gene for copper resistance. *Curr. Genet.* **9**, 133–138.

160. Henikoff, S. and Cohen, E. H. (1984). Sequences responsible for transcription termination on a gene segment in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**, 1515–1520.

161. Henikoff, S. and Furlong, C. (1983). Sequence of a *Drosophila* DNA segment that functions in *Saccharomyces cerevisiae* and its regulation by a yeast promoter. *Nucleic Acids Res.* **11**, 789–800.

162. Herrick, D., Parker, R. and Jacobson, A. (1990). Identification and comparison of stable and unstable mRNAs in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **10**, 2269–2284.

163. Heslot, H. (1990). Genetics and genetic engineering of the industrial yeast *Yarrowia lipolytica*. *Adv. Biochem. Eng. Biotechnol.* **43**, 43–73.

164. Heslot, H., Nicaud, J-M., Fabre, E., Beckerich, J.-M., Fournier, P. and Gaillardin, C. (1990). Cloning of the alkaline extracellular protease gene of *Yarrowia lipolytica* and its use to express foreign genes. In Nga, B. H. and Lee, Y. K. (Eds), *Microbiology Applications in Food Biotechnology*. Elsevier Science, Amsterdam, pp. 27–45.

165. Heyer, W. D., Sipiczki, M. and Kholi, J. (1986). Replicating plasmids in *Schizosaccharomyces pombe*: Improvement of symmetric segregation by a new genetic element. *Mol. Cell Biol.* **6**, 80–89.

166. Hildebrandt, V., Ramezani-Rad, M., Swida, U., Wrede, P., Grzesiek, S., Primke, M. and Buldt, G. (1989). Genetic transfer of the pigment bacteriorhodopsin into the eukaryote *Schizosaccharomyces pombe*. *FEBS Letts.* **243**, 137–140.

167. Hill, J., Donald, K. A. I. G. and Griffiths, D. E. (1991). DMSO-enhanced whole-cell yeast transformation. *Nucleic Acids Res.* **19**, 5791.

168. Hinnen, A., Hicks, J. B. and Fink, G. R. (1978). Transformation of yeast. *Proc. Natl. Acad. Sci. USA* **75**, 1929–1933.

169. Hinnen, A., Meyhack, B. and Heim, J. (1989). Heterologous gene expression in yeast. In Barr, P. J., Brake, A. J. and Valenzuela, P. (Eds.), *Yeast Genetic Engineering*, Butterworths, pp. 193–213.

170. Hirsch, H. H., Rendueles, P. S. and Wolf, D. H. (1989). Yeast (*Saccharomyces cerevisiae*) proteinases: structure, characteristics and function. In Walton, E. F. and Yarranton, G. T. (Eds), *Molecular and Cell Biology of Yeasts*. Blackie and Van Nostrand Reinhold, pp. 134–200.

171. Hitzeman, R. A., Chen, C. Y., Dowbenko, D. J., Renz, M. E., Lui, C., Pai, R., Simpson, N. J., Kohr, W. J., Singh, A., Chisholm, V., Hamilton, R. and Chang, C. N. (1990). Use of heterologous and homologous signal sequences for secretion of heterologous proteins from yeast. *Meth. Enzymol.* **185**, 421–440.

172. Hitzeman, R. A., Hagie, F. F., Levine, H. L., Goeddel, D. W., Ammerer, G. and Hall, B. D. (1981). Expression of human gene for interferon in yeast. *Nature* **293**, 717–723.

173. Hitzeman, R. A., Leung, D. W., Perry, L. J., Kohr, W. J., Levine, H. L. and Goeddel, D. V. (1983). Secretion of human interferons by yeast. *Science* **219**, 620–625.

174. Hodgkins, M. A., Sudbery, P. E., Kerry-Williams, S. and Goodey, A. (1990). Secretion of human serum albumin from *Hansenula polymorpha*. *Yeast* **6**, S435.

175. Hoekema, A., Kastelstein, R. A., Vasser, M. and de Boer, H. A. (1987). Codon replacement in the *PGK1* gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. *Mol. Cell. Biol.* **7**, 2914–2924.

176. Hoffman, C. S. and Winston, F. (1989). A transcriptionally regulated expression vector for the fission yeast *Schizosaccharomyces pombe*. *Gene* **84**, 473–479.

177. Holland, J. P. and Holland, M. J. (1980). Structural comparison of two nontandemly repeated yeast glyceraldehyde-3-phosphate dehydrogenase genes. *J. Biol. Chem.* **255**, 2596–2605.

178. Hollenberg, C. P. (1981). Cloning with 2-μm DNA vectors and the expression of foreign genes in *Saccharomyces cerevisiae*. *Curr. Topics Microbiol. Immunol.* **157**, 119–144.

179. Hood, M. T. and Stachow, C. (1990). Transformation of *Schizosaccharomyces pombe* by electroporation. *Nucleic Acids Res.* **18**, 688.

180. Horwitz, A. H., Chang, C. P., Better, M., Hellstrom, K. E. and Robinson, R. R. (1988). Secretion of functional antibody and Fab fragment from yeast cells. *Proc. Natl. Acad. Sci. USA* **85**, 8678–8682.

181. Hovland, P., Flick, J., Johnston, M. and Sclafani, R. A. (1989). Galactose as a gratuitous inducer of *GAL* gene expression in yeasts growing on glucose. *Gene* **83**, 57–64.

182. Hsieh, J.-H., Shih, K.-Y., Kung, H.-F., Shiang, M., Lee, L.-Y., Meng, M.-H., Chang, C.-C., Lin, H.-M., Shih, S.-C., Lee, S.-Y., Chow, T.-Y., Feng, T.-Y., Kuo, T.-T. and Choo, K.-B. (1988). Controlled fed-batch fermentation of recombinant *Saccharomyces cerevisiae* to produce hepatitis B surface antigen. *Biotechnol. Bioeng.* **32**, 334–340.

183. Huang, S., Elliott, R. C., Liu, P.-S., Koduri, R. K., Weickmann, J. L., Lee, J.-H., Blair, L. C., Ghosh-Dastida, P., Bradshaw, R. A., Bryan, K. M., Einarson, B., Kendall, R. L., Kolacz, K. H. and Saito, K. (1987). Specificity of cotranslational amino-terminal processing of proteins in yeast. *Biochemistry* **26**, 8242–8246.

184. Humphrey, T., Sadhale, P., Platt, T. and Proudfoot, N. (1991). Homologous mRNA 3' end formation in fission and budding yeast. *EMBO J.* **10**, 3503–3511.

185. Ichikawa, K., Komiya, K., Suzuki, K., Nakahara, T. and Jigami, Y. (1989). The effects of culture conditions on the secretion of human lysozyme by *Saccharomyces cerevisiae* A2-1-1A. *Agric. Biol. Chem.* **53**, 2687–2694.

186. Innis, M. A., Holland, M. J., McCabe, P. C., Cole, G. E., Wittman, V. P., Tal, R., Watt, K. W. K., Gelfand, D. H., Holland, J. P. and Meade, J. H. (1985). Expression, glycosylation and secretion of an *Aspergillus* glucoamylase by *Saccharomyces cerevisiae*. *Science* **228**, 21–26.

187. Irani, M., Taylor, W. E. and Young, E. T. (1987). Transcription of the *ADH2* gene in *Saccharomyces cerevisiae* is limited by positive factors that bind competitively to its intact promoter region on multicopy plasmids. *Mol. Cell. Biol.* **7**, 1233–1241.

188. Irniger, S., Egli, C. M. and Braus, G. H. (1991). Different classes of polyadenylation sites in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11**, 3060–3069.

189. Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983). Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**, 163–168.

190. Itoh, Y. and Fujisawa, Y. (1986). Synthesis in yeast of hepatitis B virus surface antigen modified P31 particles by gene modification. *Biochem. Biophys. Res. Commun.* **141**, 942–948.

191. Jabbar, M. A., Sivasubramanian, N. and Nayak, D. P. (1985). Influenza viral (A/WSN/33) hemagglutinin is expressed and glycosylated in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **82**, 2019–2023.

192. Jackson, M. R. and Burchell, B. (1988). Expression of human liver epoxide hydrolase in *Schizosaccharomyces pombe*. *Biochem. J.* **251**, 931–933.

193. Jacobs, E., Rutgers, T., Voet, P., Dewerchin, M., Cabezon, T. and De Wilde, M. (1989). Simultaneous synthesis and assembly of various hepatitis B surface proteins in *Saccharomyces cerevisiae*. *Gene* **80**, 279–291.

194. Jagadish, M. N., Vaughan, P. R., Irving, R. A., Azad, A. A. and Macreadie, I. G. (1990). Expression and characterisation of infectious bursal disease virus polyprotein in yeast. *Gene* **95**, 179–186.

195. James, G. and Olson, E. N. (1990). Fatty acylated proteins as components of intracellular signaling pathways. *Biochemistry* **29**, 2623–2634.

196. Janowicz, Z. A., Hening, U. and Hollenberg, C. P. (1982). Synthesis of *Escherichia coli* outer membrane protein OmpA in yeasts. *Gene* **20**, 347–358.

197. Janowicz, Z. A., Melber, K., Merckelbach, A., Jacobs, E., Harford, N., Comberbach, M. and Hollenberg, C. P. (1991). Simultaneous expression of the S and L surface antigens of hepatitis B, and formation of mixed particles in the methylotrophic yeast, *Hansenula polymorpha*. *Yeast* **7**, 431–443.

198. Janowicz, Z. A., Merckelbach, A., Eckart, M., Weydmann, U., Roggenkamp, R. and Hollenberg, C. P. (1988). Expression system based on the methylotrophic yeast *Hansenula polymorpha*. *Yeast* **4**, S155.

199. Jansen, K. U., Conroy, W. G., Claudio, T., Fox, T. D., Fujita, N., Hamill, O., Lindstrom, J. M., Luther, M., Nelson, N., Ryan, K. A., Sweet, M. T. and Hess, G. P. (1989). Expression of the four subunits of the *Torpedo californica* nicotinic acetylcholine receptor in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **264**, 15022–15027.

200. Jimenez, A. and Davies, J. (1980). Expression of a transposable antibiotic resistance element in *Saccharomyces*. *Nature* **287**, 869–871.

201. Johnston, M. (1987). A model fungal gene regulatory mechanism: the *GAL* genes of *Saccharomyces cerevisiae*. *Microbiol. Rev.* **51**, 458–476.

202. Johnston, S. A., Salmeron, J. M. and Dincher, S. S. (1987). Interaction of positive and negative regulatory proteins in the galactose-regulon of yeast. *Cell* **50**, 143–146.

203. Jones, R., Moreno, S., Nurse, P. and Jones, N. (1988). Expression of the SV40 promoter in fission yeast: identification and characterisation of an AP1-like factor. *Cell* **53**, 659–667.

204. Julius, D., Blair, L., Brake, A., Sprague, G. and Thorner, J. (1983). Yeast α-factor is processed from a larger precursor polypeptide: The essential role of a membrane-bound dipeptidyl aminopeptidase. *Cell* **32**, 839–852.

205. Kaiser, C. A., Preuss, D., Grisafi, P. and Botstein, D. (1987). Many random sequences functionally

replace the secretion signal of yeast invertase. *Science* **235**, 312–317.

206. Kamper, J., Meinhardt, F., Gunge, N. and Esser, K. (1989). *In vivo* construction of linear vectors based on killer plasmids from *Kluyveromyces lactis*: Selection of a nuclear gene results in attachment of telomeres. *Mol. Cell. Biol.* **9**, 3931–3937.
207. Kamper, K., Meinhardt, F., Gunge, N. and Esser, K. (1989). New recombinant linear DNA elements derived from *Kluyveromyces lactis* killer plasmids. *Nucleic Acids Res.* **17**, 1781.
208. Kawasaki, G. (1986). Eur. Patent Application 017114.
209. Kendall, R. L., Yamada, R. and Bradshaw, R. A. (1990). Cotranslational amino-terminal processing. *Meth. Enzymol.* **185**, 398–407.
210. Kiefhaber, T., Rudolph, R., Kohler, H.-H. and Buchner, J. (1991). Protein aggregation *in vitro* and *in vivo*: a quantitative model of the kinetic competition between folding and aggregation. *Bio/Technology* **9**, 825–829.
211. Kikuchi, Y. (1983). Yeast plasmid requires a *cis*-acting locus and two plasmid proteins for its stable maintenance. *Cell* **35**, 487–493.
212. Kikuchi, Y., Kitazawa, Y., Shimatake, H. and Yamamoto, M. (1988). The primary structure of the *leu1* gene of *Schizosaccharomyces pombe*. *Curr. Genet.* **14**, 375–379.
213. King, D. J., Walton, E. F. and Yarranton, G. T. (1989). The production of proteins and peptides from *Saccharomyces cerevisiae*. In Walton, E. F. and Yarranton, G. T. (Eds), *Molecular and Cell Biology of Yeasts*, Blackie and Van Nostrand Reinhold, pp. 107–133.
214. King, K., Dohlman, H. G., Thorner, J., Caron, M. G. and Lefkowitz, R. J. (1990). Control of yeast mating signal transduction by a mammalian β_2-adrenergic receptor and $G_s\alpha$ subunit. *Science* **250**, 121–123.
215. Kingsman, S. M., Cousens, D., Stanway, C. A., Chambers, A., Wilson, M. and Kingsman, A. J. (1990). High-efficiency yeast expression vectors based on the promoter of the phosphoglycerate kinase gene. *Meth. Enzymol.* **185**, 329–341.
216. Kingsman, S. M., Kingsman, A. J., Dobson, M. J., Mellor, J. and Roberts, N. A. (1985). Heterologous gene expression in *Saccharomyces cerevisiae*. *Biotechnol. Genet. Eng. Revs.* **3**, 377–416.
217. Kiss, G. B., Pearlman, R. E., Cornish, K. V., Friesen, J. D. and Chan, V. L. (1982). The *Herpes simplex* virus thymidine kinase is not transcribed in *Saccharomyces cerevisiae*. *J. Bacteriol.* **149** 542–547.
218. Kniskern, P. J., Hagopian, A., Montgomery, D. L., Burke, P., Dunn, N. R., Hofmann, K. J., Miller, W. J. and Ellis, R. W. (1986). Unusually high-level expression of a foreign gene (hepatitis B virus core antigen) in *Saccharomyces cerevisiae*. *Gene* **46**, 135–141.
219. Kopetzki, E., Schumacher, G. and Buckel, P. (1989). Control of formation of active soluble or inactive insoluble baker's yeast alpha-glucosidase P1 in *Escherichia coli* by induction and growth temperature. *Mol. Gen. Genet.* **216**, 149–155.
220. Kotula, L. and Curtis, J. (1991). Evaluation of foreign gene codon optimisation in yeast: expression of a mouse Ig kappa chain. *Bio/Technology* **9**, 1386–1389.
221. Kornfeld, R. and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Ann. Rev. Biochem.* **54**, 631–664.
222. Kozak, M. (1989). The scanning model for translation: an update. *J. Cell Biol.* **108**, 229–241.
223. Kramer, R. A., DeChiara, T. M., Schaber, M. D. and Hilliker, S. (1984). Regulated expression of a human interferon gene in yeast: control by phosphate concentration or temperature. *Proc. Natl. Acad. Sci. USA* **81**, 367–370.
224. Kudla, B., Persuy, M.-A., Gaillardin, C. and Heslot, H. (1988). Construction of an expression vector for the fission yeast *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **16**, 8603–8617.
225. Kukuruzinska, M. A., Bergh, M. L. E. and Jackson, B. L. (1987). Protein glycosylation in yeast. *Ann. Rev. Biochem.* **56**, 915–944.
226. Kuranda, M. J. and Robbins, P. W. (1987). Cloning and heterologous expression of glycosidase genes from *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **84**, 2585–2589.
227. Kurland, C. G. (1987). Strategies for efficiency and accuracy in gene expression 1. The major codon preference: a growth optimization strategy. *Trends Biochem. Sci.* **12**, 126–128.
228. Langridge, R., Eibel, H., Brown, J. W. S. and Feix, G. (1984). Transcription from maize storage protein gene promoters in yeast. *EMBO J.* **3**, 2467–2471.
229. Ledeboer, A. M., Edens, L., Maat, J., Visser, C., Bos, J. W., Verrips, C. T., Janowicz, Z. A., Eckart, M., Roggenkamp, R. and Hollenberg, C. P. (1985). Molecular cloning and characterisation of a gene coding for methanol oxidase in *Hansenula polymorpha*. *Nucleic Acids Res.* **13**, 3063–3082.
230. Lee, M. G. and Nurse, P. (1987). Complementation used to clone a human homologue of the fission yeast cell cycle control gene *cdc2*. *Nature* **327**, 31–35.
231. Lemontt, J. F., Wie, C.-M. and Dackowski, W. R. (1985). Expression of active human uterine tissue plasminogen activator in yeast. *DNA* **4**, 419–428.
232. Liao, X.-B., Clare, J. J. and Farabaugh, P. J. (1987). The upstream activation site of a Ty2 element of yeast is necessary but not sufficient to promote maximal transcription of the element. *Proc. Natl. Acad. Sci. USA* **84**, 8520–8524.
233. Livi, G. P., Ferrara, A., Roskin, R., Simon, P. L. and Young, P. R. (1990). Secretion of *N*-glycosylated human recombinant interleukin-1α in *Saccharomyces cerevisiae*. *Gene* **88**, 297–301.

234. Livi, G. P., Lillquist, J. S., Miles, L. M., Ferrara, A., Sathe, G. M., Simon, P. L., Meyers, C. A., Gorman, J. A. and Young, P. R. (1991). Secretion of *N*-glycosylated interleukin-1β in *Saccharomyces cerevisiae* using a leader peptide from Candida albicans. Effect of *N*-linked glycosylation on biological activity. *J. Biol. Chem.* **266**, 15348–15355.
235. Loetscher, P., Pratt, G. and Rechsteiner, M. (1991). The C terminus of mouse ornithine decarboxylase confers rapid degradation on dihydrofolate reductase. *J. Biol. Chem.* **266**, 11213–11220.
236. Loison, G., Nguyen-Juilleret, M., Alouani, F. and Marquet, M. (1986). Plasmid- transformed *URA3 FUR1* double-mutants of *S.cerevisiae*: an autoselection system applicable to the production of foreign proteins. *Bio/Technology* **4**, 433–437.
237. Loison, G., Vidal, A., Findeli, A., Roitsch, C., Balloul, J. M. and Lemoine, Y. (1989). High level of expression of a protective antigen of schistosomes in *Saccharomyces cerevisiae*. *Yeast* **5**, 497–507.
238. Lopes, T. S., Hakaart, G.-J., A. J., Koerts, B. L., Raue, H. A. and Planta, R. J. (1990). Mechanism of high-copy-number integration of pMIRY-type vectors into the ribosomal DNA of *Saccharomyces cerevisiae*. *Gene* **105**, 83–90.
239. Lopes, T. S., Klootwijk, J., Veenstra, A. E., van der Aar, P. C., van Heerikhuizen, H., Raue, H. A. and Planta, R. J. (1989). High-copy-number integration into the ribosomal DNA of *Saccharomyces cerevisiae*: a new vector for high-level expression. *Gene* **79**, 199–206.
240. Losson, R. and Lacroute, F. (1983). Plasmids carrying the yeast OMP decarboxylase structural and regulatory genes: transcriptional regulation in a foreign environment. *Cell* **32**, 371–377.
241. Losson, R., Windsor, B. and Lacroute, F. (1988). mRNA stability in *S.cerevisiae*. *Yeast* **4**, S2.
241a. Makoff, A. J. and Oxer, M. D. (1991) High level heterologous expression in *E.coli* using mutant forms of the *lac* promoter. *Nucleic Acids Res.* **19**, 2417–2421.
242. Makoff, A. J., Oxer, M. D., Romanos, M., Fairweather, N. F. and Ballantine, S. (1989). Expression of tetanus toxin fragment C in *E.coli*: high level expression by removing rare codons. *Nucleic Acids Res.* **17**, 10191–10201.
242a. Makoff, A. J., Oxer, M. D., Ballantine, S. P., Fairweather, N. F. and Charles, I. G. (1990). Protective surface antigen P69 of *Bordetella pertussis*: its characterisation and very high level expression in *Escherichia coli*. *Bio/Technology* **8**, 1030–1033.
243. Maldonado, P. and Gaillardin, C. (1972). French Patent 41913.
244. Malone, R. E. and Esposito, R. E. (1980). The *RAD52* gene is required for homothallic interconversion of mating types and spontaneous mitotic recombination in yeast. *Proc. Natl. Acad. Sci. USA* **77**, 503–507.
245. Marczynski, G. T. and Jaehning, J. A. (1985). A transcription map of a yeast centromere plasmid: unexpected transcripts and altered gene expression. *Nucleic Acids Res.* **13**, 8487–8506.
246. Marston, F. A. O. (1986). The purification of eukaryotic polypeptides synthesised in *Escherichia coli*. *Biochem J.* **240**, 1–12.
247. Matoba, S., Fukayama, J., Wing, R. and Ogrydziak, D. (1988). Intracellular precursors and secretion of the alkaline extracellular protease of *Yarrowia lipolytica*. *Mol. Cell Biol.* **8**, 4904–4916.
248. Matsumoto, K., Yoshimatsu, T. and Oshima, Y. (1983). Recessive mutations conferring resistance to carbon catabolite repression of galactokinase synthesis in *Saccharomyces cerevisiae*. *J. Bacteriol.* **153**, 14405–14414.
249. Maundrell, K. (1990). *nmt1* of fission yeast. *J. Biol. Chem.* **265**, 10857–10864.
250. McAleer, W. J., Buynak, E. B., Maigetter, R. Z., Wampler, D. E., Miller, W. J. and Hilleman, M. R. (1984). Human hepatitis B vaccine from recombinant yeast. *Nature* **307**, 178–180.
251. Mead, D. J., Gardner, D. C. J. and Oliver, S. G. (1986). The yeast 2μ plasmid: strategies for the survival of a selfish DNA. *Mol. Gen. Genet.* **205**, 417–422.
252. Meilhoc, E., Masson, J.-M. and Tessie, J. (1990). High efficiency transformation of intact yeast cells by electroporation. *Bio/Technology* **8**, 223–227.
253. Mellor, J., Dobson, M., Kingsman, A. J. and Kingsman, S. M. (1987). A transcriptional activator is located in the coding region of the yeast *PGK* gene. *Nucleic Acids Res.* **15**, 6243–6259.
254. Mellor, J., Dobson, M. J., Roberts, N. A., Kingsman, A. J. and Kingsman, S. M. (1985). Factors affecting heterologous gene expression in *Saccharomyces cerevisiae*. *Gene* **33**, 215–226.
255. Melnick, L. M., Turner, B. G., Puma, P., Price-Tillotson, B., Salvato, K. A., Dumais, D. R., Moir, D. T., Broeze, R. J. and Avgerinos, G. C. (1990). Characterisation of a nonglycosylated single chain urinary plasminogen activator secreted from yeast. *J. Biol. Chem.* **265**, 801–807.
256. Miyajima, A., Miyajima, I., Arai, K.-I. and Arai, N. (1984). Expression of plasmid R388 encoded type II dihydrofolate reductase as a dominant selectable marker in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**, 407–414.
257. Miyajima, A., Otsu, K., Schreurs, J., Bond, M. W., Abrams, J. and Arai, K. (1986). Expression of murine and human GM-CSF factors in *S. cerevisiae*: mutagenesis of the potential glycosylation sites. *EMBO J.* **5**, 1193–1197.
258. Miyamoto, C., Chizzomte, R., Crowl, R., Rupprecht, K., Kramer, R., Schaber, M., Kumar, G., Pronian, M. and Ju, G. (1985). Molecular cloning and regulated expression of the human *c-myc* gene in *Escherichia coli* and *Saccharomyces*

cerevisiae: comparison of the protein products. *Proc. Natl. Acad. Sci. USA* **82**, 7232–7236.

259. Mizunaga, T., Izawa, M., Ikeda, K. and Maruyama, Y. (1988). Secretion of an active nonglycosylated form of the repressible acid phosphatase of *Saccharomyces cerevisiae* in the presence of tunicamycin at low temperatures. *J. Biochem.* **103**, 321–326.

260. Moerschell, R. P., Hosokawa, Y., Tsunasawa, S. and Sherman, F. (1990). The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine *in vivo*. *J. Biol. Chem.* **265**, 19638–19643.

261. Moir, D. T. and Davidow, L. S. (1991). Production of proteins by secretion from yeast. *Meth. Enzymol.* **194**, 491–507.

262. Moir, D. T. and Dumais, D. R. (1987). Glycosylation and secretion of human α-1-antitrypsin by yeast. *Gene* **56**, 209–217.

263. Moreno, S., Ruiz, T., Sanchez, Y., Villanueva, J. R. and Rodriguez, L. (1985). Subcellular localisation and glycoprotein nature of the invertase from the fission yeast *Schizosaccharomyces pombe*. *Arch. Microbiol.* **142**, 370–374.

264. Murray, A. W. and Szostak, J. W. (1983). Pedigree analysis of plasmid segregation in yeast. *Cell* **34**, 911–970.

265. Murray, J. A. H. (1987). Bending the rules: the 2μ plasmid of yeast. *Mol. Microbiol.* **1**, 1–4.

266. Murray, J. A. H. and Cesarini, G. (1986). Functional analysis of the yeast plasmid partition locus STB. *EMBO J.* **5**, 3391–3399.

267. Ngsee, J. K., Hansen, W., Walter, P. and Smith, M. (1989). Cassette mutagenic analysis of the yeast invertase signal peptide: Effects on protein translocation. *Mol. Cell. Biol.* **9**, 3400–3410.

268. Ngsee, J. K. and Smith, M. (1990). Changes in a mammalian signal sequence required for efficient protein secretion by yeast. *Gene* **86**, 251–255.

269. Nicaud, J. M., Fabre, E., Beckerich, J. M., Fournier, P. and Gaillardin, C. (1989). Cloning, sequencing and amplification of the alkaline extracellular protease gene of the yeast *Yarrowia lipolytica*. *J. Biotechnol.* **12**, 285–298.

270. Nicaud, J. M., Fabre, E. and Gaillardin, C. (1989). Expression of invertase activity in *Yarrowia lipolytica* and its use as a selective marker. *Curr. Genet.* **16**, 253–260.

271. Nicaud, J. M., Fournier, P., La Bonnardiere, C., Chasles, M. and Gaillardin, C. (1991). Use of *ARS18*-based vectors to increase protein production in *Yarrowia lipolytica*. *J. Biotechnol.* **19**, 259–270.

272. Normington, K., Kohno, K., Kozutsumi, Y., Gething, M.-J. and Sambrook, J. (1989). *S. cerevisiae* encodes an essential protein homologous in sequence and function to mammalian BiP. *Cell* **57**, 1223–1236.

273. Ogrydziak, D. M. (1988). Development of genetic maps of non-conventional yeasts. *J. Basic Microbiol.* **28**, 185–196.

274. Ogrydziak, D., Bassel, J., Contopoulou, R. and Mortimer, R. (1978). Development of genetic techniques and the genetic map of the yeast *Saccharomycopsis lipolytica*. *Molec. Gen. Genet.* **163**, 229–239.

275. Ogrydziak, D. M., Demain, A. L. and Tannenbaum, S. R. (1977). Regulation of extracellular protease production in *Candida lipolytica*. *Biochem. Biophys. Acta.* **497**, 525–538.

276. Ogrydziak, D. M. and Mortimer. R. K. (1977). Genetics of extracellular protease production in *Saccharomycopsis lipolytica*. *Genetics* **87**, 621–632.

277. Orr-Weaver, T. L. and Szostak, J. W. (1983). Multiple, tandem plasmid integration in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **30**, 747–749.

278. Orr-Weaver, T. L., Szostak, J. W. and Rothstein, R. J. (1983). Genetic applications of yeast transformations with linear and gapped plasmids. *Meth. Enzymol.* **101**, 228–245.

279. Osborne, B. I. and Guarente, L. (1989). Mutational analysis of a yeast transcriptional terminator. *Proc. Natl. Acad. Sci. USA* **86**, 4097–4101.

280. Parekh, R. B., Dwek, R. A., Rudd, P. M., Thomas, J. R., Rademacher, T. W., Warren, T., Wun, T.-C., Hebert, B., Reitz, H. B., Palmier, M., Ramabhadran, T. and Tiemeier, D. C. (1989). N-glycosylation and *in vitro* enzymatic activity of recombinant human tissue plasminogen activator expressed in Chinese hamster ovary cells and a murine cell line. *Biochemistry* **28**, 7670–7679.

281. Parent, S. A., Fenimore, C. M. and Bostian, K. A. (1985). Vector systems for the expression, analysis and cloning of DNA sequences in *Saccharomyces cerevisiae*. *Yeast* **1**, 83–138.

282. Pedersen, S. (1984). *Escherichia coli* ribosomes translate *in vivo* with variable rate. *EMBO J.* **3**, 2895–2898.

283. Pelsy, F. and Lacroute, F. (1984). Effect of ochre nonsense mutations on yeast *URA1* mRNA stability. *Curr. Genet.* **8**, 277–282.

284. Pentillä, M. E., André, L., Lehtovaara, P., Bailey, M., Teeri, T. T. and Knowles, J. K. C. (1988). Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae*. *Gene* **63**, 103–112.

285. Perentesis, J. P., Genbauffe, F. S., Veldman, S. A., Galeotti, C. L., Livingston, D. M., Bodley, J. W. and Murphy, J. R. (1988). Expression of diphtheria toxin fragment A and hormone-toxin fusion proteins in toxin-resistant yeast mutants. *Proc. Natl. Acad. Sci. USA* **85**, 8386–8390.

286. Perlman, D., Halvorson, H. O. and Cannon, L. E. (1982). Presecretory and cytoplasmic invertase polypeptides encoded by distinct mRNAs derived from the same structural gene differ by a signal sequence. *Proc. Natl. Acad. Sci. USA* **79**, 781–785.

287. Picard, D., Schena, M. and Yamamoto, K. R. (1990). An inducible expression vector for both fission and budding yeast. *Gene* **86**, 257–261.

288. Pichuantes, S., Babé, L. M., Barr, P. J. and Craik, C. S. (1989). Recombinant HIV1 protease secreted by *Saccharomyces cerevisiae* correctly processes myristylated *gag* polyprotein. *Proteins: Struct. Funct. Genet.* **6**, 324–337.

289. Pobjecky, N., Rosenberg, G. H., Dinter-Gottlieb, G. and Kaufer, N. F. (1990). Expression of the β-glucuronidase gene under the control of the CaMV 35S promoter in *Schizosaccharomyces pombe*. *Molec. Gen. Genet.* **220**, 314–316.

290. Price, V., Mochizuki, D., March, C. J., Cosman, D., Deeley, M. C., Klinke, R., Clevenger, W., Gillis, S., Baker, P. and Urdal, D. (1987). Expression, purification and characterisation of recombinant murine granulocyte-macrophage colony-stimulating factor and bovine interleukin-2 from yeast. *Gene* **55**, 287–293.

291. Price, V. L., Taylor, W. E., Clevenger, Worthington, M. and Young, E. T. (1990). Expression of heterologous proteins in *Saccharomyces cerevisiae* using *ADH2* promoter. *Meth. Enzymol.* **185**, 308–318.

292. Purvis, I. J., Bettany, A. J. E., Loughlin, L. and Brown, A. J. P. (1987). Translation and stability of an *Escherichia coli* β-galactosidase mRNA expressed under the control of pyruvate kinase sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **15**, 7963–7974.

293. Purvis, I. J., Chotai, D., Dykes, C. W., Lubahn, D. B., French, F. S., Wilson, E. M. and Hobden, A. N. (1991). An androgen-inducible expression system for *Saccharomyces cerevisiae*. *Gene* **106**, 35–42.

294. Quirk, A. V., Geisow, M. J., Woodrow, J. R., Burton, S. J., Wood, P. C., Sutton, A. D., Johnson, R. A. and Dodsworth, N. (1989). Production of recombinant human serum albumin from *Saccharomyces cerevisiae*. *Biotechnol. Appl. Biochem.* **11**, 273–287.

295. Rademacher, T. W., Parekh, R. B. and Dwek, R. A. (1988). Glycobiology. *Ann. Rev. Biochem.* **57**, 785–838.

296. Reppe, S., Gabrielsen, O. S., Blingsmo, O. R., Gautvik, K. M. and Øyen, T. B. (1990). Effect of the deletion of the MFα-factor pro-region on the expression and secretion of human parathyroid hormone in *Saccharomyces cerevisiae*. *Yeast* **6**, S502.

297. Reppe, S., Gabrielsen, O. S., Olstad, O., Morrison, N., Sæther, O., Blingsmo, O. R., Gautvik, V. T., Gordeladze, J., Haflan, A. K., Voelkel, E. F., Øyen, T. B., Tashjian, A. H. Jr. and Gautvik, K. M. (1991). Characterisation of a K26Q site-directed mutant of human parathyroid hormone expressed in yeast. *J. Biol. Chem.* **266**, 14198–14201.

298. Rickard, P. A. D. and Hogan, C. B. J. (1978). Effects of glucose on the activity and synthesis of fermentative and respiratory pathways of *Saccharomyces* sp. *Biotechnol. Bioeng.* **20**, 1105–1110.

299. Riederer, M. A. and Hinnen, A. (1991). Removal of N-glycosylation sites of the yeast acid phosphatase severely affects protein folding. *J. Bact.* **173**, 3539–3546.

300. Rinas, U., Rise, B., Jaenicke, R., Broker, M., Karges, H. E., Kupper, H. A. and Zettlmeissi, G. (1990). Characterisation of recombinant factor XIIIa produced in *Saccharomyces cerevisiae*. *Bio/Technology* **8**, 543–545.

301. Rogers, S., Wells, R. and Rechsteiner, M. (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **234**, 364–368.

302. Roggenkamp, R., Hansen, H., Eckart, M., Janowicz, Z. and Hollenberg, C. P. (1986). Transformation of the methylotrophic yeast *Hansenula polymorpha* by autonomous replication and integration vectors. *Mol. Gen. Genet.* **202**, 302–308.

303. Roggenkamp, R., Kustermann-Kuhn, B. and Hollenberg, C. P. (1981). Expression and processing of bacterial β-lactamase in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **78**, 4466–4470.

304. Romanos, M. A., Beesley, K. M. and Clare, J. J. (1991). Direct selection of stabilised yeast *URA3* transformants with 5-fluorouracil. *Nucleic Acids Res.* **19**, 187.

305. Romanos, M. A. and Boyd, A. (1988). A transcriptional barrier to expression of cloned toxin genes of the linear plasmid k1 of *Kluyveromyces lactis*: evidence that native k1 has novel promoters. *Nucleic Acids Res.* **16**, 7333–7350.

306. Romanos, M. A., Clare, J. J., Beesley, K. M., Rayment, F. B., Ballantine, S. P., Makoff, A. J., Dougan, G., Fairweather, N. F. and Charles, I. G. (1991). Recombinant *Bordetella pertussis* pertactin (P69) from the yeast *Pichia pastoris*: high-level production and immunological properties. *Vaccine* **9**, 901–906.

307. Romanos, M. A., Makoff, A. J., Fairweather, N. F., Beesley, K. M., Slater, D. E., Rayment, F. B., Payne, M. M. and Clare, J. J. (1991). Expression of tetanus toxin fragment C in yeast: gene synthesis is required to eliminate fortuitous polyadenylation sites in AT-rich DNA. *Nucleic Acids Res.* **19**, 1461–1467.

308. Rose, M., Grisafi, P. and Botstein, D. (1984). Structure and function of the yeast *URA3* gene: expression in *Escherichia coli*. *Gene* **29**, 113–124.

309. Rose, M. D., Misra, L. M. and Vogel, J. P. (1989). *KAR2*, a karyogamy gene, is the yeast homolog of the mammalian BiP/GRP78 gene. *Cell* **57**, 1211–1221.

310. Rosenberg, S., Barr, P. J., Najarian, R. C. and Hallewell, R. A. (1984). Synthesis in yeast of a

functional oxidation resistant mutant of human α_1-antitrypsin. *Nature* **312**, 77–80.

311. Rosenberg, S., Coit, D. and Tekamp-Olson, P. (1990). Glyceraldehyde-3-phosphate dehydrogenase-derived expression cassettes for constitutive synthesis of heterologous proteins. *Meth. Enzymol.* **185**, 341–351.

312. Rothblatt, J. A., Webb, J. R., Ammerer, G. and Meyer, D. I. (1987). Secretion in yeast: structural features influencing the post-translational translocation of prepro-α-factor *in vitro*. *EMBO J.* **6**, 3455–3463.

313. Rothstein, R. J. (1983). One-step gene disruption in yeast. *Meth. Enzymol.* **101**, 202–211.

314. Rothstein, S. J., Lahners, K. N., Lazarus, C. M., Baulcombe, D. C. and Gatenby, A. A. (1987). Synthesis and secretion of wheat α-amylase in *Saccharomyces cerevisiae*. *Gene* **55**, 353–356.

315. Rudolph, H. K., Antebi, A., Fink, G. R., Buckley, C. M., Dorman, T. E., LeVitre, J., Davidow, L. S., Mao, J. and Moir, D. T. (1989). The yeast secretory pathway is perturbed by mutations in *PMR1*, a member of a Ca^{2+} ATPase family. *Cell* **58**, 133–145.

316. Ruohonen, L., Hackman, P., Lehtovaara, P., Knowles, J. K. C. and Keränen, S. (1987). Efficient secretion of *Bacillus amyloliquefaciens* α-amylase by its own signal peptide from *Saccharomyces cerevisiae* host cells. *Gene* **59**, 161–170.

317. Russell, P. R. (1983). Evolutionary divergence of the mRNA transcription initiation mechanism in yeast. *Nature* **301**, 167–169.

318. Russell, P. R. (1985). Transcription of the triosephosphate isomerase gene of *Schizosaccharomyces pombe* initiates from a start point different from that in *Saccharomyces cerevisiae*. *Gene* **40**, 125–130.

319. Russell, P. (1989). Gene cloning and expression in fission yeast. In Nasim, A., Young, P. and Johnson, B. F. (Eds), *Molecular Biology of the Fission Yeast*. Academic Press, New York, pp. 243–271.

320. Russell, P. R. and Hall, B. D. (1983). The primary structure of the alcohol dehydrogenase gene from the fission yeast *Schizosaccharomyces pombe*. *J. Biol. Chem.* **258**, 143–149.

321. Sabin, E. A., Lee-Ng, C. T., Shuster, J. R. and Barr, P. J. (1989). High-level expression and *in vivo* processing of chimeric ubiquitin fusion proteins in *Saccharomyces cerevisiae*. *Bio/Technology* **7**, 705–709.

322. Sakai, A., Ozawa, F., Higashizaki, T., Shimizu, Y. and Hishinuma, F. (1991). Enhanced secretion of human nerve growth factor from *Saccharomyces cerevisiae* using an advanced δ-integration system. *Bio/Technology* **9**, 1382–1385.

323. Sakai, K., Sakaguchi, J. and Yamamoto, M. (1984). High frequency cotransformation by copolymerisation of plasmids in the fission yeast *Schizosaccharomyces pombe*. *Mol. Cell Biol.* **4**, 651–656.

324. Saliola, M., Shuster, J. R. and Falcone, C. (1990). The alcohol dehydrogenase system in the yeast, *Kluyveromyces lactis*. *Yeast* **6**, 193–204.

325. Sambucetti, L. C., Schaber, M., Kramer, R., Crowl, R. and Curran, T. (1986). The *fos* gene product undergoes extensive post-translational modification in eukaryotic but not in prokaryotic cells. *Gene* **43**, 69–77.

326. Santangelo, G. M., Tornow, J., McLaughlin, C. S. and Moldave, K. (1988). Properties of promoters cloned randomly from the *Saccharomyces cerevisiae* genome. *Mol. Cell. Biol.* **8**, 4217–4224.

327. Santiago, T. C., Bettany, A. J. E., Purvis, I. J. and Brown, A. J. P. (1987). Messenger RNA stability in *Saccharomyces cerevisiae*: The influence of translation and poly(A) tail length. *Nucleic Acids Res.* **15**, 2417–2429.

328. Sato, T., Uemura, H., Izumoto, Y., Nakao, J., Nakamura, Y. and Matsubara, K. (1989). The conformation of mature α-amylase conditions its secretion from yeast. *Gene* **83**, 355–365.

329. Sauer, N., Freidlander, K. and Graml-Wicke, U. (1990). Primary structure, genomic organisation and heterologous expression of a glucose transporter from *Arabidopsis thaliana*. *EMBO J.* **9**, 3045–3050.

330. Schafer, W. R., Trueblood, C. E., Yang, C.-C., Mayer, M. P., Rosenberg, S., Poulter, C. D., Kim, S.-H. and Rine, J. (1990). Enzymatic coupling of cholesterol intermediates to a mating pheromone precursor and to the ras protein. *Science* **249**, 1133–1139.

331. Schein, C. H. and Noteborn, M. H. M. (1988). Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperature. *Bio/Technology* **6**, 291–294.

332. Schena, M., Picard, D. and Yamamoto, K. R. (1991). Vectors for constitutive and inducible gene expression in yeast. *Meth. Enzymol.* **194**, 389–398.

333. Schonberger, O., Hirst, T. R. and Pines, O. (1991). Targeting and assembly of an oligomeric bacterial enterotoxoid in the endoplasmic reticulum of *Saccharomyces cerevisiae*. *Mol. Microbiol.* **5**, 2663–2671.

334. Schultz, L. D., Hofmann, K. J., Mylin, L. M., Montgomery, D. L., Ellis, R. W. and Hopper, J. E. (1987). Regulated overproduction of the *GAL4* gene product greatly increases expression from galactose-inducible promoters on multi-copy expression vectors in yeast. *Gene* **61**, 123–133.

335. Schultz, L. D., Tanner, J., Hofmann, K. J., Emini, E. A., Condra, J. H., Jones, R. E., Kieff, E. and Ellis, R. W. (1987). Expression and secretion in yeast of a 400-kDa envelope glycoprotein derived from Epstein-Barr virus. *Gene* **54**, 113–123.

336. Scorer, C. A., Carrier, M. J. and Rosenberger, R. F. (1991). Amino acid misincorporation during high-level expression of mouse epidermal growth

factor in *Escherichia coli*. *Nucleic Acids Res.* **19**, 3511–3516.

337. Shah, D. N., Purohit, A. P. and Sriprakash, R. S. (1982). Preliminary studies on a citric acid producing strain of *Saccharomycopsis lipolytica*. *Enzyme Microbiol. Technol.* **4**, 69–83.

338. Sharp, P. M. and Cowe, E. (1991). Synonymous codon usage in *Saccharomyces cerevisiae*. *Yeast* **7**, 657–678.

339. Shaw, G. and Kamen, R. (1986). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**, 659–667.

340. Shen, H.-S., Bastien, L., Nguyen, T., Fung, M. and Slilaty, S. N. (1989). Synthesis and secretion of hepatitis middle surface antigen by the methylotrophic yeast *Hansenula polymorpha*. *Gene* **84**, 303–309.

341. Sherwood, R. (1991). Protein fusions: bioseparation and application. *Trends Biotechnol.* **9**, 1–3.

342. Shortle, D., Novick, P. and Botstein, D. (1984). Construction and genetic characterisation of temperature-sensitive mutant alleles of the yeast actin gene. *Proc. Natl. Acad. Sci. USA* **81**, 4889–4893.

343. Shuster, J. R. (1989). Regulated transcriptional systems for the production of proteins in yeast: regulation by carbon source. In Barr, P. J., Brake, A. J. and Valenzuela, P. (Eds), *Yeast Genetic Engineering*. Butterworths, pp.83–108.

344. Shuster, J. R. (1990). *Kluyveromyces lactis* glyceraldehyde-3-phosphate dehydrogenase and alcohol dehydrogenase-1 genes are linked and divergently transcribed. *Nucleic Acids Res.* **18**, 4271.

345. Shuster, J. R., Lee, H. and Moyer, D. L. (1990). Integration and amplification of DNA at yeast delta sequences. *Yeast* **6**, S79.

346. Shuster, J. R., Moyer, D. and Irvine, B. (1987). Sequence of the *Kluyveromyces lactis URA3* gene. *Nucleic Acids Res.* **15**, 8573.

347. Shuster, J. R., Moyer, D. L., Lee, H., Dennis, A., Smith, B. and Merryweather, J. P. (1989). Yeast mutants conferring resistance to toxic effects of cloned human insulin-like growth factor I. *Gene* **83**, 47–55.

348. Sierkstra, L. N., Verkabel, J. M. A. and Verrips, C. T. (1991). Optimisation of a host/vector system for heterologous gene expression by *Hansenula polymorpha*. *Curr. Genet.* **19**, 81–87.

349. Silve, S., Monod, M., Hinnen, A. and Haguenauer-Tsapis, R. (1987). The yeast acid phosphatase can enter the secretory pathway without its N-terminal signal sequence. *Mol. Cell. Biol.* **7**, 3306–3314.

350. Singh, A., Lugovoy, J. M., Kohr, W. J. and Perry, L. J. (1984). Synthesis, secretion and processing of α-factor-interferon fusion proteins in yeast. *Nucleic Acids Res.* **12**, 8927–8938.

351. Skinner, R. H., Bradley, S., Brown, A. L., Johnson, N. J. E., Rhodes, S., Stammers, D. K. and Lowe, P. N. (1991). Use of the Glu-Glu-Phe C-terminal epitope for rapid purification of the catalytic domain of normal and mutant *ras* GTPase-activating proteins. *J. Biol. Chem.* **266**, 14163–14166.

352. Sledziewski, A. Z., Bell, A., Kelsay, K. and MacKay, V. L. (1988). Construction of temperature-regulated yeast promoters using the *MATα2* repression system. *Bio/Technology* **6**, 411–416.

353. Sledziewski, A. Z., Bell, A., Yip, C., Kelsay, K., Grant, F. J. and MacKay, V. L. (1990). Superimposition of temperature regulation on yeast promoters. *Meth. Enzymol.* **185**, 351–366.

354. Sleep, D., Belfield, G. P., Ballance, D. J., Steven, J., Jones, S., Evans, L. R., Moir, P. D. and Goodey, A. R. (1991). *Saccharomyces cerevisiae* strains that overexpress heterologous proteins. *Bio/Technology* **9**, 183–187.

355. Sleep, D., Belfield, G. P. and Goodey, A. R. (1990). The secretion of human serum albumin from the yeast *Saccharomyces cerevisiae* using five different leader sequences. *Bio/Technology* **8**, 42–46.

356. Sleep, D., Ogden, J. E., Roberts, N. A. and Goodey, A. R. (1991). Cloning and characterisation of the *Saccharomyces cerevisiae* glycerol-3-phosphate dehydrogenase (*GUT2*) promoter. *Gene* **101**, 89–96.

357. Smith, R. A., Duncan, M. J. and Moir, D. T. (1985). Heterologous protein secretion from yeast. *Science* **230**, 1219–1224.

358. Søgaard, M. and Svensson, B. (1990). Expression of cDNAs encoding barley α-amylase 1 and 2 in yeast and characterisation of the secreted proteins. *Gene* **94**, 173–179.

359. Sreekrishna, K., Nelles, L., Potenz, R., Cruze, J., Mazzaferro, P., Fish, W., Motohiro, F., Holden, K., Phelps, D., Wood, P. and Parker, K. (1989). High-level expression, purification, and characterisation of recombinant human tumour necrosis factor synthesised in the methylotrophic yeast *Pichia pastoris*. *Biochemistry* **28**, 4117–4125.

360. Sreekrishna, K., Potenz, R. B., Cruze, J. A., McCombie, W. R., Parker, K. A., Nelles, L., Mazzaferro, P. K., Holden, K. A., Harrison, R. G., Wood, P. J., Phelps, D. A., Hubbard, C. E. and Fuke, M. (1988). High level expression of heterologous proteins in methylotrophic yeast *Pichia pastoris*. *J. Basic Microbiol.* **28**, 265–278.

361. Sreekrishna, K., Tschopp, J. F. and Fuke, M. (1987). Invertase gene (*SUC2*) of *Saccharomyces cerevisiae* as a dominant marker for transformation of *Pichia pastoris*. *Gene* **59**, 115–125.

362. Sreekrishna, K., Webster, T. D. and Dickson, R. C. (1984). Transformation of *Kluyveromyces lactis* with the kanamycin (G418) resistance gene of Tn *903*. *Gene* **28**, 73–81.

363. St. John, T. P. and Davis, R. W. (1981). The organisation and transcription of the galactose gene

cluster of *Saccharomyces*. *J. Mol. Biol.* **152**, 285–315.
364. Stark, M. J. R., Boyd, A., Mileham, A. J. and Romanos, M. A. (1990). The plasmid-encoded killer system of *Kluyveromyces lactis*: a review. *Yeast* **6**, 1–29.
365. Stark, M. J. R. and Milner, J. S. (1988). Cloning and analysis of the *Kluyveromyces lactis TRP1* gene: a chromosomal locus flanked by genes encoding inorganic pyrophosphatase and histone H3. *Yeast* **5**, 35–50.
366. Stepien, P. P., Brousseau, R., Wu, R., Narang, S. and Thomas, D. Y. (1983). Synthesis of a human insulin gene VI. Expression of the synthetic proinsulin gene in yeast. *Gene* **24**, 289–297.
367. Steube, K., Chaudhuri, B., Märki, W., Merryweather, J. P. and Heim, J. (1991). α-factor-leader-directed secretion of recombinant human-insulin-like growth factor I from *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **198**, 651–657.
368. Strasser, A. W., Selk, R., Dohmen, R. J., Nierman, T., Bielefeld, M., Seeboth, P., Tu, G. and Hollenberg, C. P. (1989). Analysis of the α-amylase gene of *Schwanniomyces occidentalis* and the secretion of its gene product in transformants of different yeast genera. *Eur. J. Biochem.* **184**, 699–706.
369. Struhl, K. (1989). Molecular mechanisms of transcriptional regulation in yeast. *Ann. Rev. Biochem.* **58**, 1051–1077.
370. Struhl, K. and Davis, R. W. (1980). A physical, genetic, and transcriptional map of the yeast *his3* gene of *Saccharomyces cerevisiae*. *J. Mol. Biol.* **136**, 309–332.
371. Sutton, A. and Broach, J. R. (1985). Signals for transcription initiation and termination in the *Saccharomyces cerevisiae* plasmid 2μm circle. *Mol. Cell. Biol.* **5**, 2770–2780.
372. Takabayashi, K., Märki, W., Wolf, D. H. and Heim, J. (1990). Degradation of yeast recombinant desulfato-hirudin variant 1 and its mutants by the two carboxypeptidases yscY and ysca. *Yeast* **6**, S490.
373. Tammi, M., Ballou, L., Taylor, A. and Ballou, C. E. (1987). Effect of glycosylation on yeast invertase oligomer stability. *J. Biol. Chem.* **262**, 4395–4401.
374. Tanguy-Rougeau, C., Chen, X. J., Wesolowski-Louvel, M. and Fukuhara, H. (1990). Expression of a foreign Km[R] gene in linear killer DNA plasmids in yeast. *Gene* **91**, 43–50.
375. Thill, G. P, Davis, G. R., Stillman, C., Holtz, G., Brierley, R., Engel, M., Buckholtz, R., Kinney, J., Provow, S., Vedvick, T. and Seigel, R. S. (1990). Positive and negative effects of multi-copy integrated expression vectors on protein expression in *Pichia pastoris*. In Heslot, H., Davies, J., Florent, J., Bobichon, L., Durand, G. and Penasse, L. (Eds), *Proceedings of the 6th International Symposium on Genetics of Microorganisms*, Vol. II. Societé Française de Microbiologie, Paris, France, pp. 477–490.
376. Thim, L., Hansen, M. T., Norris, K., Hoegh, I., Boel, E., Forstrom, J., Ammerer, G. and Fiil, N. P. (1986). Secretion and processing of insulin precursors in yeast. *Proc. Natl. Acad. Sci. USA* **83**, 6766–6770.
377. Tobe, S., Takami, T., Ikeda, S. and Mitsugi, K. (1976). Production and some enzymatic properties of alkaline proteinase of *Candida lipolytica*. *Agric. Biol. Chem.* **40**, 1087–1092.
378. Tottrup, H. V. and Carlsen, S. (1990). A process for the production of human proinsulin in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **35**, 339–348.
379. Toyama, R. and Okayama, H. (1990). Human chorionic gonadotropin α and human cytomegalovirus promoters are extremely active in the fission yeast *Schizosaccharomyces pombe*. *FEBS Letts.* **268**, 217–221.
379a. Trimble, R. B., Atkinson, P. H., Tschopp, J. F., Townsend, R. R. and Maley, F. (1991). Structure of oligosaccharides on *Saccharomyces SUC2* invertase secreted by the methylotrophic yeast *Pichia pastoris*. *J. Biol. Chem.* **266**, 22807–22817.
380. Tschopp, J. F., Brust, P. F., Cregg, J. M., Stillman, C. A. and Gingeras, T. R. (1987). Expression of the *lacZ* gene from two methanol-regulated promoters in *Pichia pastoris*. *Nucleic Acids Res.* **15**, 3859–3876.
381. Tschopp, J. F., Sverlow, G., Kosson, R., Craig, W. and Grinna, L. (1987). High-level secretion of glycosylated invertase in the methylotrophic yeast, *Pichia pastoris*. *Bio/Technology* **5**, 1305–1308.
382. Tschumper, G. and Carbon, J. (1980). Sequence of a yeast DNA fragment containing a chromosomal replicator and the *TRP1* gene. *Gene* **10**, 157–166.
383. Tuite, M. F., Dobson, M. J., Roberts, N. A., King, R. M., Burke, D. C., Kingsman, S. M. and Kingsman, A. J. (1982). Regulated high efficiency expression of human interferon-alpha in *Saccharomyces cerevisiae*. *EMBO J.* **1**, 603–608.
384. Turner, B. G., Avgerinos, G. C., Melnick, L. M. and Moir, D. T. (1991). Optimisation of pro-urokinase secretion from recombinant *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **37**, 869–875.
385. Urdea, M. S., Merryweather, J. P., Mullenbach, D. C., Corr, D., Heberlein, U., Valenzuela, P. and Barr, P. J. (1983). Chemical synthesis of a gene for human epidermal growth factor urogastrone and its expression in yeast. *Proc. Natl. Acad. Sci. USA* **80**, 7461–7465.
386. Utsumi, S., Kanamori, J., Kim, C.-S., Sato, T. and Kito, M. (1991). Properties and distribution of soybean proglycinin expressed in *Saccharomyces cerevisiae*. *J. Agric. Food Chem.* **39**, 1179–1186.
387. Valenzuela, P., Medina, A., Rutter, W. J., Ammerer, G. and Hall, B. D. (1982). Synthesis

and assembly of hepatitis B virus surface antigen particles in yeast. *Nature* **298**, 347–350.
388. Van Arsdell, J. N., Kwok, S., Schweickart, V. L., Ladner, M. B., Gelfand, D. H. and Innis, M. A. (1987). Cloning, characterisation and expression in *Saccharomyces cerevisiae* of endoglucanase I from *Trichoderma reesei*. *Bio/Technology* **5**, 60–64.
389. van den Heuvel, J. J., Bergkamp, R. J. M., Planta, R. J. and Raue, H. A. (1989). Effect of deletions in the 5'-noncoding region on the translational efficiency of phosphoglycerate kinase mRNA in yeast. *Gene* **79**, 83–95.
390. van den Heuvel, J. J., Planta, R. J. and Raue, H. A. (1990). Effect of leader primary structure on the translational efficiency of phosphoglycerate kinase mRNA in yeast. *Yeast* **6**, 473–482.
391. van den Berg, J. A., van der Laken, K. J., van Ooyen, A. J. J., Renniers, T. C. H. M., Reitveld, K., Schaap, A., Brake, A. J., Bishop, R. J., Schultz, K., Moyer, D., Richman, M. and Schuster, J. R. (1990). *Kluyveromyces* as a host for heterologous gene expression: Expression and secretion of prochymosin. *Bio/Technology* **8**, 135–139.
392. van Dijl, J. M., de Jong, A., Smith, H., Bron, S. and Venema, G. (1991). Signal peptidase I overproduction results in increased efficiencies of export and maturation of hybrid secretory proteins in *Escherichia coli*. *Mol. Gen. Genet.* **227**, 40–48.
393. Varshavsky, A., Bachmair, A., Finley, D., Gonda, D. K. and Wunning, I. (1989). Targeting of proteins for degradation. In Barr, P. J., Brake, A. J. and Valenzuela, P. (Eds), *Yeast Genetic Engineering*, Butterworths, pp. 109–143.
394. Vogel, K. and Hinnen, A. (1990). The yeast phosphatase system. *Mol. Microbiol.* **4**, 2013–2017.
395. Volkert, F. C., Wilson, D. W. and Broach, J. R. (1989). Deoxyribonucleic acid plasmids in yeasts. *Microbiol. Revs.* **53**, 299–317.
396. Wagenbach, M., O'Rourke, K., Vitez, L., Wieczorek, A., Hoffman, S., Durfee, S., Tedesco, J. and Stetler, G. (1991). Synthesis of wild type and mutant hemoglobins in *Saccharomyces cerevisiae*. *Bio/Technology* **9**, 57–61.
397. Walmsley, R. M., Gardner, D. C. and Oliver, S. G. (1983). Stability of a cloned gene in yeast grown in chemostat culture. *Mol. Gen. Genet.* **194**, 361–365.
398. Walton, E. F. and Yarranton, G. T. (1989). Negative regulation of gene expression by mating type. In Walton, E. F. and Yarranton, G. T. (Eds), *Molecular and Cell Biology of Yeasts*. Blackie and Van Nostrand Reinhold, pp. 43–69.
399. Wampler, D. E., Lehman, E. D., Boger, J., McAleer, W. J. and Scolnick, E. M. (1985). Multiple chemical forms of hepatitis B surface antigen produced in yeast. *Proc. Natl. Acad. Sci. USA* **82**, 6830–6834.
400. Webster, T. D. and Dickson, R. C. (1983). Direct selection of *Saccharomyces cerevisiae* resistant to the antibiotic G418 following transformation with a DNA vector carrying the kanamycin resistance gene of Tn903. *Gene* **26**, 243–252.
401. Wegner, E. H., (1983). Biochemical conversions by yeast fermentation at high-cell densities. U.S. patent 4414329.
402. Wesolowski-Louvel, M., Prior, C., Mamessier, P., Goffrini, P., Ferrero. I., Sor, F. and Fukuhara, H. (1990). *Kluyveromyces lactis* and its plasmids. In Heslot, H., Davies, J., Florent, J., Bobichon, L., Durand, G. and Penasse, L. (Eds). *Proceedings of the 6th International Symposium on Genetics of Microorganisms*, Vol. II. Societé Francaise de Microbiologie, Paris, France, pp. 519–532.
403. Wickerham, L. J., Kurtzman, C. P. and Herman, A. I. (1970). Sexual reproduction in *Candida lipolytica*. *Science* **167**, 1141.
404. Wilson, W., Braddock, M., Adams, S., Rathjen, P., Kingsman, S. and Kingsman, A. (1988). HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. *Cell* **55**, 1159–1169.
405. Wing, R. A. and Ogrydziak, D. M. (1985). Development of the genetics of the dimorphic yeast *Yarrowia lipolytica*. In Timberlake, W. E. (Ed), *Molecular Genetics of Filamentous Fungi*. A.R. Liss, Inc., New York, pp. 367–381.
406. Wood, C. R., Boss, M. A., Kenton, J. M., Calvert, J. E., Roberts, N. A. and Emtage, J. S. (1985). The synthesis and *in vivo* assembly of functional antibodies in yeast. *Nature* **314**, 446–449.
407. Xuan, J.-W., Fournier, P. and Gaillardin, C. (1988). Cloning of the *LYS5* gene encoding saccharopine dehydrogenase from the yeast *Yarrowia lipolytica* by targeted integration. *Curr. Genet.* **14**, 15–21.
407a. YaDeau, J. T., Klein, C. and Blobel, G. (1991). Yeast signal peptidase contains a glycoprotein and the *SecII* gene product. *Proc. Natl. Acad. Sci. USA* **88**, 517–521.
408. Yamada, T. and Ogrydziak, D. M. (1983). Extracellular acid proteases produced by *Saccharomycopsis lipolytica*. *J. Bacteriol.* **154**, 23–31.
409. Yamamoto, M. (1989). Fission yeast. In, Barr, P. J., Brake, A. J. and Valenzuela, P. (Eds), *Yeast Genetic Engineering*. Butterworths, pp. 53–64.
410. Yamashita, T., Tonouchi, N., Vozumi, T. and Beppu, T. (1987). Secretion of Mucor rennin, a fungal aspartic protease of *Mucor pusillus* by recombinant yeast cells. *Mol. Gen. Genet.* **210**, 462–467.
411. Yeh, P., Fleer, R., Maury, I. and Mayaux, J.-F. (1990). Secretion of naturally N-glycosylated human proteins in *Kluyveromyces lactis*. Abstract D14, 6th International Symposium on Genetics of Industrial Organisms, Strasbourg.
412. Yellen, G. and Migeon, J. C. (1990). Expression of *Torpedo* nicotinic acetylcholine receptor subunits in yeast is enhanced by the use of yeast signal sequences. *Gene* **86**, 145–152.

413. Yocum, R. R., Hanley, S., West, R. and Ptashne, M. (1984). Use of *lacZ* fusions to delimit regulatory elements of the inducible *GAL1-GAL10* promoter in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4**, 1985–1998.

414. Zaman, Z., Brown, A. J. P. and Dawes, I. W. (1992). A 3' transcriptional enhancer within the coding sequence of a yeast gene encoding the common subunit of two multienzyme complexes. *Mol. Microbiol.* **6**, in press.

415. Zaret, K. S. and Sherman, F. (1982). DNA sequence required for efficient transcription termination in yeast. *Cell* **28**, 563–573.

416. Zaret, K. S. and Sherman, F. (1984). Mutationally altered 3' ends of yeast *CYC1* mRNA affect transcript stability and translational efficiency. *J. Mol. Biol.* **177**, 107–136.

417. Zaworski, P. G., Marotti, K. R., MacKay, V., Yip, C. and Gill, G. S. (1989). Production and secretion of porcine urokinase in *Saccharomyces cerevisiae*: characterisation of the secreted gene product. *Gene* **85**, 545–551.

418. Zealey, G. R., Goodey, A. R., Piggott, J. R., Watson, M. E., Cafferkey, R. C., Doel, S. M., Carter, B. L. A. and Wheals, A. E. (1988). Amplification of plasmid copy number by thymidine kinase expression in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **211**, 155–159.

419. Zhu, X. L., Ward, C. and Weissbach, A. (1984). Control of *Herpes simplex* virus thymidine kinase gene expression in *Saccharomyces cerevisiae* by a yeast promoter sequence. *Mol. Gen. Genet.* **194**, 31–41.

420. Zsebo, K. M., Lu, H.-S., Fieschko, J. C., Goldstein, L., Davis, J., Duker, K., Suggs, S. V., Lai, P.-H. and Bitter, G. A. (1986). Protein secretion from *Saccharomyces cerevisiae* directed by the prepro-α-factor region. *J. Biol. Chem.* **261**, 5858–5865.

Critical Reviews in

Plant Sciences

B. V. Conger
Editor

Volume 15 / Issue 1 / 1996

Inhibition of mRNA levels and activities by *trans*-cinnamic acid in elicitor-induced bean cells, *Phytochemistry*, 27, 2109, 1988.

170. **Mavandad, M., Edwards, R., Liang, X., Lamb, C. J., and Dixon, R. A.**, Effects of *trans*-cinnamic acid on expression of the bean phenylalanine ammonia-lyase gene family, *Plant Physiol.*, 94, 671, 1990.

171. **Loake, G. J., Choudhary, A. D., Harrison, M. J., Mavandad, M., Lamb, C. J., and Dixon, R. A.**, Phenylpropanoid pathway intermediates regulate transient expression of a chalcone synthase gene promoter, *Plant Cell*, 3, 829, 1991.

172. **Loake, G. J., Faktor, O., Lamb, C. J., and Dixon, R. A.**, Combination of H-box [CCTACC(N)$_7$CT] and G-box (CACGTG) cis elements is necessary for feed-forward stimulation of a chalcone synthase promoter by the phenylpropanoid-pathway intermediate *p*-coumaric acid, *Proc. Natl. Acad. Sci. U.S.A.*, 89, 9230, 1992.

173. **Durst, F., Benveniste, I., Salaun, J.-P., and Werck-Reichhart, D.**, Function, mechanism and regulation of cytochrome P-450 enzymes in plants, *Biochem. Soc. Trans.*, 20, 353, 1992.

Critical Reviews in Plant Sciences, 15(4):285–423 (1996)

Mechanisms of Intracellular Protein Transport and Targeting in Plant Cells

Allison R. Kermode
Simon Fraser University, Department of Biological Sciences, Burnaby, B.C., Canada V5A 1S6

Referee: Roger Beachy, Cell Biology–MRC-7, Scripps Research Inst., 10666 N. Torrey Pines Rd., La Jolla, CA 92037

TABLE OF CONTENTS

I. INTRODUCTION 287
II. PATHWAYS OF PROTEIN TRANSPORT IN PLANT CELLS 288
III. GENERAL PRINCIPLES OF PROTEIN TARGETING 289
IV. GENERAL METHODOLOGY TO STUDY PROTEIN TARGETING 290
V. THE BIOSYNTHETIC (SECRETORY) TRANSPORT PATHWAY 293
A. Unique Features of the Plant ER and its Relationship to the Golgi Complex 294
B. Signal-Dependent and Signal-Independent Steps Along the Secretory Pathway 295
1. Evidence for Default Pathway 295
2. Signal-Dependent Steps Along the Secretory Pathway 298
a. Signal-Dependent Translocation Across the ER Membrane 299
b. Signal-Dependent Retention of Soluble Proteins in the ER Lumen 305
c. Targeting Signals for Membrane Proteins of the Secretory Pathway 309
i. Topology of Integral Membrane Proteins 309
ii. Localization Signals of Membrane Proteins 310
C. Membrane Growth and Lipid Transfer 313
D. Cotranslational and Posttranslational Modifications Along the Secretory Pathway 314
1. Glycosylation 314
2. Folding and Oligomer Assembly 319
3. Formation of Disulfide Bonds 324
4. Amino Acid Modifications 325
5. Proteolytic Processing 326
E. Transport of Proteins Through the ER and Golgi Complex 327
1. Vesicles Involved in Transport 329
2. Mechanisms of Vesicle Budding, Targeting, and Fusion: Cytosolic Factors, Receptors, and GTP-Binding Proteins 331
F. Hormonal and Environmental Controls Over Protein Secretion 341
G. Signal-Mediated Sorting of Proteins to Vacuoles and Retention in ER-Derived Protein Bodies 345
1. Mechanisms of Storage Protein Deposition in Seeds 345
2. Biogenesis of ER-Derived Protein Bodies and Mechanisms of Protein Retention 346

0735-2689/96/$.50

3. Evidence for an Alternative Routing Mechanism to Vacuole-Derived Protein Bodies in Cereal Seeds that Bypasses the Golgi Complex 349
4. Expression of Storage Proteins in Heterologous Hosts: Universality of Targeting Machinery 350
5. Vacuolar Targeting in Yeast: Genetic and Biochemical Studies 351
6. Identifying Plant Vacuolar Targeting Signals 354
a. Use of Transgenic Yeast as the Heterologous Host System 354
b. Use of Transgenic Plants as the Heterologous Host System 356
c. Targeting of Membrane Proteins to the Tonoplast 359
d. Other Components of the Plant Vacuolar Targeting Machinery: The Search for Receptors 360
VI. APPLICATIONS TO GENETIC ENGINEERING: MAXIMIZING LEVELS OF GENE EXPRESSION IN FOREIGN PLANT HOSTS 361
A. Enhancing Protein Stability in Transgenic Host Plants 362
B. Enhancing Posttranslational Processes: Folding and Oligomer Assembly 364
VII. TARGETING OF PROTEINS TO THE NUCLEUS 365
A. Nuclear Structure and Components 365
B. Mechanisms of Nuclear Targeting 366
1. Macromolecular Transport Across the Nuclear Envelope 366
2. Protein Import in Animal and Yeast Cells 367
3. Nuclear Import and Export of Proteins that Undergo Shuttling 369
4. Nuclear Transport in Plants 370
a. Nuclear Localization of Gene Regulatory Proteins 371
b. Nuclear Localization of Plant Viral Proteins 372
c. Nuclear Localization of Nucleic Acids during *Agrobacterium*-Mediated Plant Transformation 372
VIII. PROTEIN TARGETING INTO PEROXISOMES AND GLYOXYSOMES 372
IX. PROTEIN TARGETING INTO AND WITHIN CHLOROPLASTS 376
A. Protein Targeting to the Stroma 376
B. Protein Targeting Toward the Thylakoid Membrane System 379
C. Protein Targeting to the Envelope Membranes and Intermembrane Space 381
D. Energy Requirements 381
E. Precursor Binding and the Identification of Receptors and Accessory Proteins 382
F. Assembly of Oligomeric Protein Complexes: Role of Chaperonins 384
X. PROTEIN TARGETING INTO AND WITHIN MITOCHONDRIA 384
A. Protein Targeting to the Matrix 385
B. Protein Targeting to the Inner Membrane and Intermembrane Space 386
C. Other Components Involved in Protein Import and Assembly 387
XI. PROSPECTS FOR ENGINEERING CHLOROPLAST AND MITOCHONDRIAL PROTEINS 390
A. Herbicide Tolerance 390
B. Increasing Photosynthetic Productivity 391
C. Allotopic Expression of Mitochondrial Genes 392
XII. SUMMARY 392

ABSTRACT: The specificity of protein targeting processes is the basis of maintaining structural and functional integrity of the cell, enabling the various subcellular compartments to carry out their unique metabolic roles. Studies in plants have progressed markedly in the last 5 years, and many of the specific signals involved in the transport and targeting of proteins to the nucleus, chloroplast, mitochondrion and microbody, and to organelles along the secretory pathway (endoplasmic reticulum [ER], Golgi complex, and vacuole) have been characterized. Exciting prospects include the identification of receptors involved in the recognition of protein targeting signals, mechanisms of vesicle targeting, and the role of mRNA targeting. Although important exceptions exist, a striking feature of the mechanisms and cellular machinery of protein targeting is their universality — among plants, animals, and eukaryotic microorganisms — and even between prokaryotes and eukaryotes. More information is required about the structural features of proteins that allow for their stable accumulation in a particular subcellular compartment, of particular interest to the plant genetic engineer. Our understanding of the rules that govern protein folding and oligomer assembly and how these processes relate to a protein's ultimate stability in the cell is limited.

KEY WORDS: protein targeting, protein folding, protein assembly, protein stability, secretory pathway, molecular chaperone.

I. INTRODUCTION

A recent focus of considerable study in the field of plant cell biology is the elucidation of the molecular mechanisms involved in the intracellular transport and targeting of proteins from their site of synthesis to their final destination. The eukaryotic cell must utilize mechanisms to ensure the faithful expression of genetic information at the level of functionally and correctly localized proteins. These posttranslational controls may be imposed at several steps between synthesis of proteins on cytosolic (or organellar) polyribosomes and their final localization and functional configuration within the cell. With the exception of a few proteins of chloroplasts and mitochondria, all proteins are synthesized on free or bound polyribosomes of the cytosol, and so must be transported to, or across, one or more membranes to reach their final subcellular location. Some very fundamental questions concerning these regulatory events are being addressed in plants. For example, what is the nature of the signal and sorting machinery that target proteins to different compartments of the cell? What are the mechanisms of signal "decoding" and of protein insertion into, or translocation across, the target membrane? What are the posttranslational modifications involved in protein maturity and the acquisition of a stable three-dimensional conformation? Also, what determines the half-life of a protein in a particular cell type? The maintenance of correct intracellular targeting of proteins is likely to be particularly challenging in plants in which there is an additional organelle — the chloroplast — requiring extensive protein traffic from the cytosol and a high degree of cytoarchitecture to carry out its functions.[1] But the importance of the specificity of these processes to the cell cannot be underestimated because it is the very basis of maintaining structural and functional integrity, enabling the various compartments of the cell to carry out their diverse metabolic roles.

The focus of this review is on the various mechanisms involved in correct protein targeting in plant cells as well as some of the experimental approaches used to elucidate these events. Recent information on the mechanisms of protein targeting into and within the chloroplast, mitochondrion, peroxisome, glyoxysome, and nucleus will be covered as well as the targeting of proteins to organelles of the secretory system. Some examination of the components involved in the posttranslational modifications of proteins as well as the structure/architecture of the secretory pathway are reviewed; these topics have also been covered in recent reviews.[2-4]

Much progress has been made in recent years in the area of protein transport and targeting in plant cells. However, many of the mechanistic details concerning related events (e.g., protein translocation and vesicle targeting) remain to be elucidated. Thus, where appropriate, reference is made to analogous mechanisms in other eukaryotic organisms, with the awareness that these "heterologous systems" may not fully reflect the processes in plants.

Plant genetic engineers have endeavored to manipulate metabolic pathways to improve plant productivity and to engineer seed crops with

greater nutritional value or protein content. With respect to the latter, this has taken the form of attempts to enhance gene transcription and mRNA stability, or translatability. More recently, a novel strategy has been toward enhancing protein stability and the "mis- or "retargeting" of proteins into different subcellular compartments. Some of these studies are discussed.

II. PATHWAYS OF PROTEIN TRANSPORT IN PLANT CELLS

Major targets of protein transport in plant cells include the chloroplast, the mitochondrion, the nucleus, and the peroxisome/glyoxysome (Figure 1).[5] Proteins destined for these organelles are translated on "free" polyribosomes of the cytosol; concurrently, or shortly thereafter, they are imported into the appropriate organelle by direct recognition of specific targeting signals. In contrast, transport to the vacuole and cell surface occurs along the endomembrane system (the endoplasmic reticulum [ER], Golgi apparatus, and transport vesicles); proteins destined for this so-called "secretory" pathway are synthesized on polyribosomes associated with the ER and enter a common "gateway" by first translocating across the membrane of the ER into the lumen (Figure 1). The organelles of the secretory pathway are functionally connected by vesicular traffic; proteins move along this intercommunicating system in the lumen or in the membrane of vesicles that form from one compartment and fuse to the next compartment. The major route of transport is from the ER to the Golgi complex and from there to the vacuole or to the cell surface (plasma membrane, cell wall, or extracellular space). Thus, the protein composition of the ER and Golgi complex includes both permanent residents as well as proteins *en route* to their final destinations.

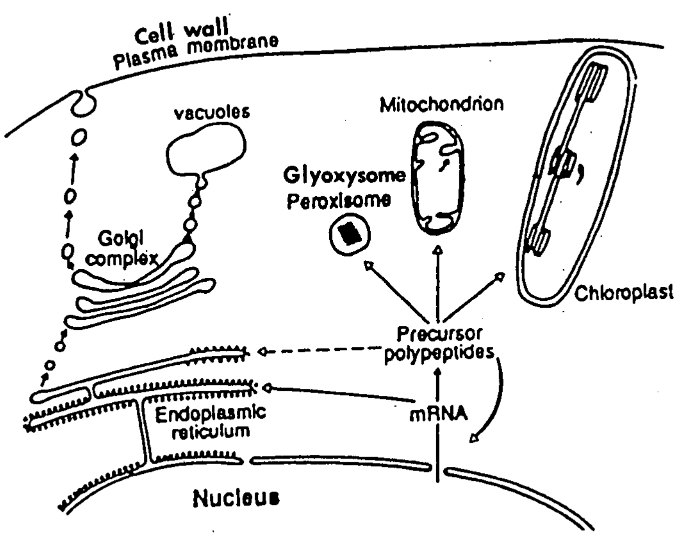


FIGURE 1. The major pathways of protein transport within the plant cell. Based on Verner, K. and Shatz, G., *Science*, 241, 1307, 1988. (With permission from the American Association for the Advancement of Science.)

III. GENERAL PRINCIPLES OF PROTEIN TARGETING

Our understanding of the mechanisms of protein targeting is far from complete, particularly within plants, but present knowledge suggests that some general principles apply:

1. Targeting information resides in the protein itself. It may be in the form of a discrete signal (e.g., a specific stretch of contiguous amino acids located internally or at the extreme end of the protein) (Figure 2A).[6] Often only the general features of targeting signals are conserved across a group of proteins targeted to the same organelle (i.e., there is no strict primary consensus signal). Thus, the secondary or tertiary structural features of the signal may be recognized by the targeting machinery. Discrete signals are also often context dependent, that is, the topogenic information must be accessible to the targeting machinery involved in its recognition. Less common than discrete signals are topogenic sequences referred to as signal patches; these targeting signals are formed from noncontiguous regions or domains of the polypeptide chain that are brought together in the folded protein (i.e., they are conformation-dependent) (Figure 2B). A single protein may have multiple targeting signals. In some cases these are distinct. However, they can be overlapping and, moreover, their functions may be affected by covalent modifications such as phosphorylation.[7] Quite different signals (and transport pathways) may serve to target proteins to the same organelle, an example being those responsible for directing proteins to plant vacuoles.
2. Several factors may influence the accuracy and efficiency of protein targeting. For example, the localization of mRNAs to specific subcellular regions or compartments

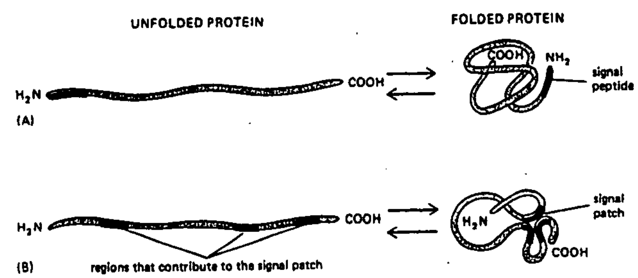


FIGURE 2. Two ways that targeting signals can be organized in proteins. (A) The signal is in a single discrete stretch of amino acid sequence that is exposed in the folded protein. These signals often occur at the end of the polypeptide chain (as shown), but they can also be located elsewhere. They are normally detected experimentally by their effect on the intracellular sorting of other proteins when they are attached to them by recombinant DNA methods (see Figure 3). (B) A signal patch can be formed by the juxtaposition of amino acids from regions that are physically separated before the protein folds (as shown); alternatively, separate "patches" on the surface of the folded protein that are spaced a fixed distance apart could form the signal. In either case, the transport signal depends on the three-dimensional conformation of the protein. For this reason it is very difficult to locate this type of signal precisely. (From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D., *Molecular Biology of the Cell*, 2nd ed., Garland Publishing, New York, 1989. With permission from Garland Publishing.)

provides an important mechanism for targeting proteins to the sites where their activity is required.

3. Targeting sequences are probably recognized and "decoded" by specific protein receptors; interactions with lipids of the membrane bilayer may also be important, perhaps playing a facilitory role in the initial binding process to receptors on the target membrane.
4. The types of signals ("topogenic" sequences) involved in protein targeting are limited and can be classified on the basis of the nature of the target membrane and, in some cases, on the mechanism of membrane translocation or integration (e.g., involvement of specific protein effectors). Examples are signal sequences, stop-transfer sequences, retention sequences, and sorting sequences.[8] Not all membranes are translocation-competent and most translocate only specific proteins.[5] **Signal** sequences initiate translocation of proteins across specific cellular membranes; they interact with protein receptors/translocators, which effect unidirectional translocation. **Stop-transfer** sequences interrupt the translocation process that was previously initiated by a signal peptide and yield integration of proteins into translocation-competent membranes. Multiple start-and-stop transfer sequences can result in integral membrane proteins that span the bilayer multiple times.[6,9-11] **Retention** sequences function to retain proteins in certain compartments, such as the ER. **Sorting** sequences act as determinants for post-translocational traffic of subpopulations of proteins, originating in translocation-competent donor membranes (and compartments) and going to translocation-incompetent receiver membranes (compartments). For example, sorting signals direct proteins along branched pathways such as that encountered at the *trans*-Golgi network (TGN), where proteins have different transport fates after having shared a common route through the ER and Golgi complex (Figure 1; see later discussion). **Insertion** sequences may initiate unilateral integration of proteins into the lipid bilayer without the mediation of a distinct protein effector.[8] Some examples of specific targeting signals of eukaryotic proteins that fall into the major categories of topogenic sequences are presented in Table 1.[10]
5. The mechanism of protein translocation across the target membrane may involve interactions of a variety of cytosolic and membrane components; it often requires an unfolded or "loosely folded" conformation of the protein undergoing translocation, the maintenance of which may involve assistance by another protein.
6. In most cases, but by no means all, the targeting signal is transient and cleaved following protein translocation; however, this step is generally not an obligatory component of the targeting/translocation process per se.
7. Proteins fold into their final three-dimensional structure (and in some cases, assemble into oligomers) shortly after translocation is completed; often these processes are assisted by a general group of proteins termed "molecular chaperones."
8. Although important exceptions exist, a striking feature of our understanding of the mechanisms of protein targeting is their universality — among animals, plants, and eukaryotic microorganisms and even between prokaryotes and eukaryotes.

IV. GENERAL METHODOLOGY TO STUDY PROTEIN TARGETING

Various methods have evolved over the years to study protein targeting in eukaryotic cells. Current approaches commonly involve the application of recombinant DNA techniques to engineer genes (e.g., to effect specific mutations or deletions in a given gene, or to construct chimeric genes). Transfer and expression of these modified genes in host cells or organisms allow the subsequent analysis of protein targeting in a heterologous environment (e.g., the subcellular localization of the genetically engineered proteins). In addition, where isolation and preservation of in-

TABLE 1
Examples of Targeting Sequences on Eukaryotic Proteins

General type of topogenic sequence	Targeting event	Specific signal
Signal sequences	Translocation into ER lumen (signal peptide)	+ 0 0 0 0 + MKWVTFLLLLFISGSAFS^RGVF
	Chloroplast (stroma) targeting (transit peptide)	00 00 +00 0++0 0 0 0 + + MAPAVMASSATTVAPFQGLKSTAGLPVSRRSGSLGSVSNGGRIRC^M-
Stop-transfer sequences	ER stop-transfer sequence	+00 0 + + -KSSIASPFFIIGLIIGLFLVLRVGIH-
Retention sequences	ER retention sequence (soluble protein)	+-- -KDEL
	ER retention sequence (membrane protein)	++0 --++ -RRSPFIDEKKMP
Sorting sequences	Plant cell vacuole	+0+ + 0 -HSRFNPIRLPT
	Yeast vacuole	+ -LQRP-
	Animal cell lysosome	Mannose-6-phosphate "tag"

Note: Sequences are written in the single letter code. Charged residues are indicated + or –, and amino acids with hydroxyl groups on side chains (Thr or Ser residues) are indicated 0. ^ marks the position of endopeptidase cleavage.

Based on Austen, B. M. and Westwood, O. M. R., *Protein Targeting and Secretion*, Oxford University Press, Oxford, 1991. With permission from Oxford University Press.

tact organelles is feasible (e.g., chloroplasts, mitochondria), protein targeting can be studied via *in vitro* systems. In animals and yeast, systems for reconstituted *in vitro* protein transport between two organelles (e.g., the ER and Golgi complex) have also been developed.

Because the targeting signal is part of the protein itself, engineering of the respective gene involves modifications to the coding region only (Figure 3).[12] Two general approaches have commonly been undertaken:

1. A **"loss of function" approach** (Figure 3A) involves an analysis of the fate of mutants in which gene sequences (presumed to encode targeting signals of the corresponding protein) are deleted. The assumption here is that deletions in a targeting domain will cause the mutant protein to bypass the normal sorting reaction and, hence, result in its mislocalization in the transgenic host cells. If a targeting domain is implicated by these studies, it may be further delimited or defined by mutational analysis of specific residues. This "loss of function" approach has certain limitations because abolished targeting may be attributed to the deletion causing a nonspecific protein conformational change (affecting transport), rather than to the removal of a sequence involved in targeting per se. Hence, this approach is often combined with another one.
2. In the **"gain-of-function" approach** (Figure 3B), chimeric (fusion) genes are constructed, consisting of various lengths of coding sequence (from the gene coding the targeted protein) linked to an intact reporter gene coding sequence. This "gain-of-function" approach allows an identification of protein targeting domains that are sufficient to direct the reporter protein into the organelle under study. For example, Figure 3B shows this approach applied to identify the sorting signal of a vacuolar protein. One advantage of gene fusions is that the "reporter" or "passenger" protein may be easier to detect and localize in the heterologous host cells than the targeted protein itself; ideally, the added sequences present in the chimeric protein should not affect (i.e., reduce) reporter protein detection (e.g., enzymatic activity).[13] Cytosolic proteins are often preferred as reporters because they should not contain targeting signals; however, a prudent choice of a reporter protein is necessary because they may be unable to cross certain membranes.[14] This approach has been very useful for identifying discrete targeting signals on proteins, that is, those consisting of a specific stretch of contiguous amino acids. However, it is of limited use for the delineation of topogenic sequences referred to as "signal patches" (see earlier), those made up of noncontiguous regions of the polypeptide chain (Figure 2).

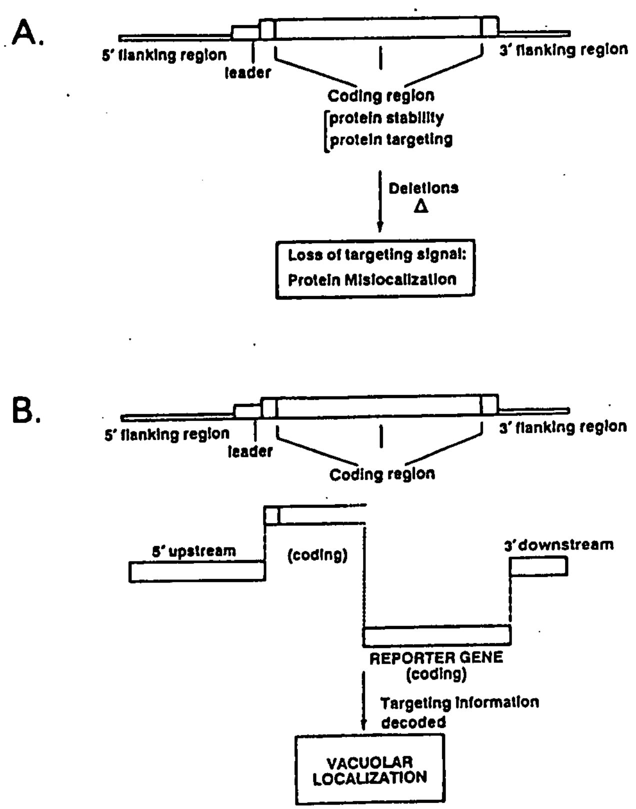


FIGURE 3. Approaches to study protein targeting. (A) "Loss of function" approach. Deletions in a targeting domain result in mislocalization of the mutant protein in the heterologous host cells. (B) "Gain of function" approach. The figure shows the approach applied to identify the amino acid sequences (or protein domain) sufficient to direct a reporter protein to the target organelle (e.g., the vacuole). Refer to text for further explanation. (From Kermode, A. R., in *Mechanisms of Plant Growth and Improved Productivity: Modern Approaches*, Basra, A. S., Ed., Marcel Dekker, New York, 1994, 317. With permission from Marcel Dekker.)

Specific methods have also been developed to identify the receptors involved in the recognition and decoding of specific sorting signals, as well as other components of the transport and targeting machinery; these are discussed in subsequent sections.

V. THE BIOSYNTHETIC TRANSPORT PATHWAY

As mentioned previously, some proteins synthesized in the cytosol (e.g., those subsequently routed to the vacuole, or destined for secretion from the cell) are transported along the biosynthetic transport pathway[15,16] or "secretory" pathway (reviewed in References 4 and 17). The organelles of this pathway include the ER (rough and smooth), Golgi apparatus, TGN, endosomes (where the exocytic and endocytic pathways converge), a variety of transport vesicles, vacuoles, the tonoplast, and the plasma membrane. As in other eukaryotic cells, the protein composition of these organelles in plant cells is highly dependent on the tissue or developmental stage of the cell (i.e., strict temporal and spatial regulation of gene expression is operative); it may also be influenced by external environmental signals. For example, vacuoles are acidic compartments that perform a variety of functions dependent upon the physiological status of the plant cell. Within the reserve tissues of many seeds, they perform a dual function — that is, as temporary storage depots (during seed development, when large amounts of storage proteins and other specialized proteins are synthesized and accumulated), and as sites of macromolecular hydrolysis, during the postgerminative phase of the plant lifecycle, when seedling growth must be supported. It is during this latter phase that the vacuole accumulates several hydrolytic enzymes involved in reserve mobilization, and the plant cell vacuole is thought of as equivalent to the animal lysosome. In contrast to the numerous small storage vacuoles found in the reserve tissues of seeds, developing cells of vegetative tissues (e.g., mesophyll cells of leaves) typically have one (or more) very prominent vacuole occupying much of the cell's volume. These vacuoles, like others, are multifunctional; in addi-

tion to providing a driving force for cell growth by maintaining turgor, they also have an important lytic role.[18]

Many vegetative reserve tissues (e.g., tubers, roots, bark) accumulate storage proteins and lectins in their vacuoles (often in a seasonal-dependent manner);[19–23] other vegetative tissues accumulate vacuolar storage proteins only when subjected to various stresses (e.g., nitrogen stress).[24] Pathogen invasion induces the synthesis of a variety of enzymes (including β-glucanases and chitinases) that are subsequently directed to both vacuolar and extracellular locations.[25,26]

Other major plant cell proteins transported along the "secretory" pathway are those destined for the tonoplast (the vacuolar membrane), the plasma membrane, and extracellular matrix (cell wall). Proteins transported to the plasma membrane include those that are essential for cellulose synthesis, hormone perception, and ion transport as well as those that mediate interactions with the cell wall and/or cytoskeleton.[27] Included in the most abundant proteins transported to the cell wall are the principal structural cell wall proteins, the hydroxyproline-rich glycoproteins (HRGPs) (e.g., the extensins), and the glycine- and proline-rich proteins, as well as some arabinogalactan proteins (found in specialized mucilages and gums) and the solanaceous lectins.[27–30] Other less abundant cell wall proteins include the cysteine-rich thionins, 28- and 70-kDa water-regulated proteins, a histidine-tryptophan-rich protein, and several cell wall enzymes (peroxidases, phosphatases, invertases, proteases).[30,31] The cell wall also appears to be a highly dynamic compartment; the abundance of several of its composite proteins is sensitive to external cues (such as wounding and pathogen invasion), as well as being subject to the normal spatial/temporal controls (Table 2).[30,32–48]

Tissues in the cereal caryopsis (i.e., the aleurone layer and scutellum) have a glandular function (during postgerminative seedling growth), and secrete a large number of hydrolytic enzymes that play an important role in endosperm storage reserve mobilization. Aleurone layers isolated from mature cereal grains have provided an important system for studying protein transport and secretion; upon treatment with gibberellin (GA) (and in the presence of calcium), a large proportion of this tissue's protein synthetic capacity becomes devoted to the production of several extracellular enzymes, including α-amylase, β-glucanase, ribonuclease, xylosidase, and acid phosphatase, each of which is actively secreted.[29,49]

The membranes surrounding organelles of the secretory pathway also represent specialized compartments for the many reactions catalyzed by integral membrane proteins. Often overlooked is the fact that different plant cell types contain different proportions of various membranes (reviewed in Reference 50), and this has raised questions about the mechanisms controlling membrane growth and changes in membrane lipid content/composition requiring the movement of lipids between organelles (see later discussion).

A. Unique Features of the Plant ER and Its Relationship to the Golgi Complex

The plant ER is a complex network of cisternae and tubules containing a single internal space (ER lumen). It is a dynamic organelle changing in organization during differentiation or environmental stress. It typically represents a fairly large proportion of a plant cell's total membrane (e.g., 30%), being particularly prominent in cell types exhibiting a high ratio of protein secretion or protein transport to the vacuole (Figure 4). The plant ER, in contrast to that of other eukaryotic cells, has several unique functions. For example, in seeds of some plant species, the ER is a site of aggregation and accumulation of some classes of storage proteins. Other more specialized roles of the plant ER may include anchoring the cytoskeleton, communication between the exterior of the cell and the cytoplasm, and communication between contiguous cells of the plant body. The former two roles have been attributed to the cortical ER of plant cells (i.e., the endomembrane system located in the peripheral layer of cytoplasm adjacent to the plasma membrane) (reviewed in Reference 51). A functional dynamic endomembrane continuum also exists between contiguous plant cells.[52] Intercellular communication between plant cells occurs through plasmodesmata, tubular structures embedded in the plant cell wall in association with the plasma membrane and ER. The ER associated with these plasmodesmata forms a communication pathway permitting intercellular passage of lipids, fatty acids, and important signal transduction molecules (e.g., the lipid secondary message, diacylglycerol). It may also provide the basis for plant hormone signal transmission, viral propagation, and pathogenic responses.[52]

TABLE 2
Examples of Plant Proteins Transported Along the Secretory Pathway and Conditions Regulating Their Expression

Protein class	Conditions
Structural cell[a] wall proteins	
Extensins (dicot)	Wounding, fungal infection, viral infection, fungal elicitors, endogenous elicitors, ethylene, red light, heat shock, gravity, glutathione, cell culturing, development
GRPs (dicot)	Development, viral infection, salicylic acid, abscisic acid, drought stress, wounding
GRPs (monocot)	Development, water stress, abscisic acid, mercuric chloride, wounding
PRPs	Wounding, endogenous elicitors, fungal elicitor, ethylene, cell culturing, light, red light, development
PRPs (nodulins)	Development
AGPs	Development, wounding
Secreted enzymes[b]	
α-Amylase	Abscisic acid, gibberellic acid, calcium, water stress, desiccation/rehydration
Chitinases	Development, fungal elicitors, abiotic elicitors
Plasma membrane proteins[c]	
pp34 (defense-related)	Oligogalacturonide elicitors
Plasma membrane ATPase	Development, hormones
Tonoplast/vacuolar proteins[d]	
TIP (tonoplast intrinsic proteins)	Development, water stress/osmoregulation
Slow vacuolar ion channel (of tonoplast)	Abscisic acid, gibberellic acid, calcium
Seed storage proteins (e.g., napin)	Development, abscisic acid, jasmonic acid, sugars, desiccation/rehydration
Vegetative storage proteins	Development (temporal), water deficit, wounding, nitrogen supply, jasmonic acid

[a] Reviewed in Showalter;[30] see references therein.
[b] Jones et al.;[34] Bol et al.;[25] Van den Bulcke et al.;[26] Kermode.[35,36]
[c] Jacinto et al.;[37] Michelet et al.[38]
[d] Maurel et al.;[39] Johnson et al.;[40] Bethke and Jones;[41] DeLisle and Crouch;[42] Wilen et al.;[43,44] Jiang et al.;[45] Reinbothe et al.;[46] Staswick;[24] Mason and Mullet;[47] Mason et al.[48]

The use of vital dyes has revealed some features of the internal organization of the ER. For example, the fluorochrome 3,3-dihexyloxacarbocyanine iodide stain shows that the ER of onion bulb epidermal cells is subdivided into three domains: cisternae (putative rough ER) and two forms of smooth ER, a peripheral tubular network and long tubular strands.[53]

The Golgi complex of plant cells (unlike that of most other eukaryotic cells) is actively engaged in the biosynthesis of the polysaccharide components of the cell wall (e.g., pectins and hemicelluloses), and many plant cells secrete far more polysaccharide than glycoprotein. In different cell types there may be distinct variations in the structural relations between the ER and Golgi stacks, perhaps dependent upon the nature of the secretory product. Not surprisingly, a closer association of the Golgi complex with the ER occurs in cell types exhibiting a higher rate of protein secretion than in plant cells engaged primarily in the secretion of carbohydrate-based molecules, in which only remote relations are observed (reviewed in Reference 3).

B. Signal-Dependent and Signal-Independent Steps Along the Secretory Pathway

1. Evidence for a Default Pathway

Proteins destined for transport along the secretory pathway are synthesized on ribosomes as-

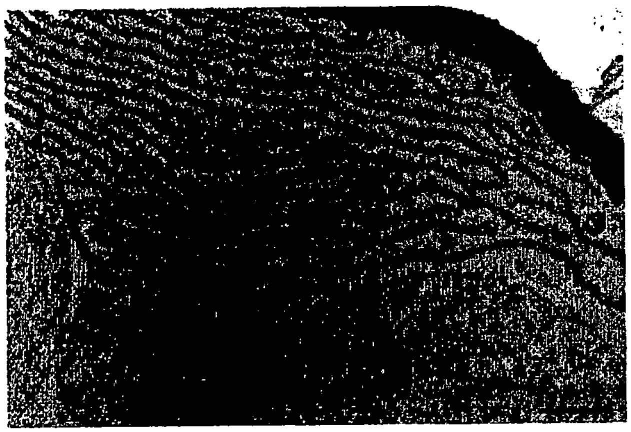


FIGURE 4. Plant endoplasmic reticulum (ER). Proliferation of the ER in a plant cell specialized for the accumulation of high amounts of protein in vacuoles. The electron micrograph shows a portion of a cell of developing common bean cotyledons. The cotyledons are synthesizing large amounts of storage protein that is cotranslationally translocated into the ER lumen for transport along the secretory pathway to protein storage vacuoles. The dark-stained area is the cell wall. Portions of protein storage vacuoles are visible at the bottom and left. Bar = 1 μm.(From Vitale, A., Ceriotti, A., and Denecke, J., *J. Exp. Bot.*, 44, 1417, 1993. With permission from Oxford University Press. Courtesy of Franco Faoro.)

sociated with the ER. The first step of entry into the pathway is mediated by a N-terminal signal peptide on the nascent protein that directs a transmembrane translocation from the cytosol to the lumen of the ER,[54] a step generally accompanied by signal peptide cleavage (see later discussion). The ER contains a mixture of proteins with multiple destinations. Some proteins will become permanent residents of the ER; others are exported to the Golgi apparatus for retention there, or for subsequent distribution to the cell surface or lysosome/vacuole. How this complex traffic pattern is organized, and what steps along the pathway occur by default (i.e., are signal-independent), or alternatively, require specific targeting signals, have been subjects for both speculation and debate (reviewed in References 16 and 55 through 60). For example, there have been arguments for proteins requiring specific signals for export out of the ER; others have argued for nonselective export out of the ER, with positive signals being required for ER retention. The available evidence from eukaryotic systems (including plants) indicates that each of the steps along the intracellular route leading to constitutive secretion from the cell (i.e., ER to Golgi complex, movement through the Golgi stacks, and movement from the Golgi complex to the cell surface) occurs by default: secreted and plasma membrane proteins are carried along by a nonselective "bulk-flow" process.[16,17,59,61] Proteins destined for targets other than the cell surface must contain additional positive topogenic information. Thus, additional information is required for selective retention in the ER (or the Golgi complex), or for diversion of proteins away from the bulk-flow pathway to one of the branch pathways within the TGN (e.g., to the animal lysosome or plant/yeast vacuole) (Figure 5).

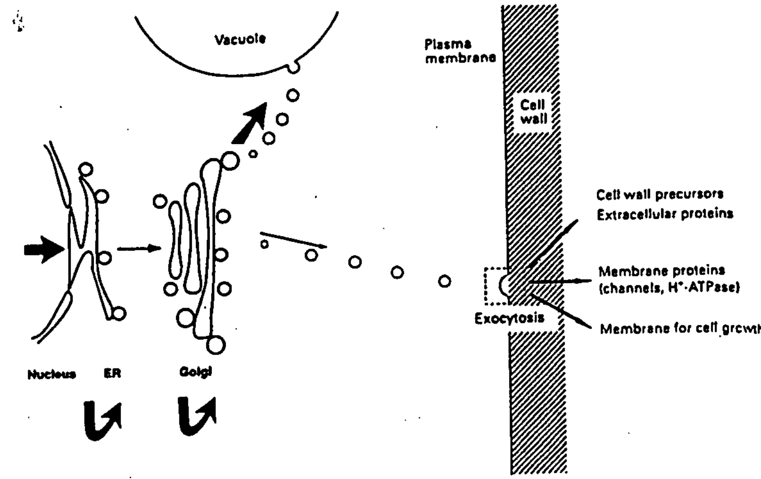


FIGURE 5. Current model of protein transport through the plant secretory pathway. Thin arrows depict signal-independent transport and represent bulk-flow (the "default" pathway). Thick arrows denote transport steps that are signal-mediated, including translocation across the ER membrane, retention of resident proteins in the ER and Golgi complex, and transport to vacuoles.

There are several arguments in favor of a bulk-flow model in which (constitutive) secretion is the default fate for a soluble protein containing no additional specific targeting information (over that of the signal peptide).[17,29,62–65] Perhaps the most striking evidence for the existence of a default pathway in animal cells comes from the work of Wieland et al.,[66] who devised a bulk phase marker of the secretory pathway. A synthetic tripeptide consisting of the glycosylation sequence Asn-Tyr-Thr, esterified to make it membrane permeable, can enter cells, reach the ER lumen where it becomes glycosylated (thus rendering it membrane impermeant), and then be rapidly secreted, predominantly along the normal pathway (i.e., via the Golgi complex). It is assumed that the tripeptide contains no transport or retention signals and defines the bulk-flow rate of ER to cell surface movement and thereby the default pathway.[57,61] Moreover, the bulk-flow rate defined by this marker is sufficiently rapid to account for the export of most secretory or membrane proteins, being close to the rate of transport of the most rapidly secreted proteins.[66] Although there are some valid criticisms regarding the appropriateness of the glycosylated tripeptide as a bulk phase marker,[55,58,59] there is currently no compelling reason to postulate the existence of a positive ER export signal for any protein.[59]

The loss or disruption of a defined positive sorting signal on a protein (e.g., by mutation or deletion) generally leads to its transport along the default pathway (i.e., missorting and secretion). Similarly, when signal-dependent sorting to the animal lysosome or to specialized secretory vesicles (involved in regulated secretion) is saturated or incapacitated, the bulk-flow pathway to the cell surface seems to be the route taken by default. In I-cells that are unable to tag lysosomal hydrolases with the required sorting signals (i.e., mannose-6-phosphate residues) and in cells deficient in one of the mannose-6-phosphate receptors (see later discussion), a large fraction of lysosomal hydrolases are constitutively secreted.[67,68]

Significant mistargeting and secretion also occurs in yeast cells when two proteins (normally targeted to the vacuole) are overexpressed, presumably because the capacity of vacuolar sorting machinery is surpassed.[69,70] Removal of putative ER-retention signals from ER-resident proteins causes them to be transported along the secretory pathway, even though they would not do so normally, and hence, are not expected to have an export/transport signal.[70–72] Conversely, mutations in secreted or plasma membrane proteins that abolish transport and lead to their retention have subsequently been shown to affect transport out of the ER by indirect means (reviewed in References 60, 73, and 74). Deletion or disruption of a specific export signal is not operative here; rather, these proteins are unable to undergo the normal folding/assembly processes required for subsequent transport out of the ER (see later discussion).

Disruption of signal-dependent sorting in the TGN by chemical treatments that affect the pH of the transport pathway (and presumably receptor-ligand interactions such as chloroquine and monensin) also cause mistargeting and secretion.[67,75–77] Mistargeting and transport of proteins along the bulk-flow pathway occur when animal cells are treated with the sodium ionophore monensin, which raises the pH of the TGN.[77] A similar phenomenon occurs in plant cells; treatment of pea and jackbean cotyledons results in the transport of storage proteins and lectins (normally accumulated in the storage vacuole/protein body) to the plasma membrane and cell wall.[78,79]

Likewise, in roots of transgenic tomato plants, proteinase inhibitors I and II (primarily vacuolar in location) are found in greater abundance at the cell wall following monensin treatment.[80] Not all vacuolar storage proteins behave in this manner; pro-barley lectin and bean phytohemagglutinin (PHA) are retained intracellularly following treatment with monensin.[81,82] The varying effects of this ionophore on the trafficking of secretory products may reflect structural and functional cell type-specific differences in the organization of plant Golgi stacks.[83]

Secretion is the outcome in plant cells when a protein contains no additional targeting information over that of the signal peptide.[62–65] In studies of chimeric proteins expressed in transgenic tobacco, a signal peptide (e.g., from the vacuolar proteins PHA and patatin) is sufficient to direct a cytosolic reporter protein (e.g., pea albumin 2 and β-glucuronidase, or GUS) into the ER and leads to transport along the secretory pathway via a default/bulk-flow mechanism. Likewise, in tobacco cells transformed by electroporation, bacterial cytosolic enzymes, utilized as reporters, enter the secretory pathway (when linked to the signal peptide of an extracellular pathogenesis-related protein, PR1, or that of the cell wall glycoprotein, extensin) and are secreted.[64,84] Secretion in these transformed tobacco cells is relatively inefficient; however, the results are indicative that secretion can occur independently of active sorting by non-specific migration through the secretory pathway.

2. Signal-Dependent Steps Along the Secretory Pathway

In contrast to the signal-independent export of proteins out of the ER (and intra-Golgi transport by bulk flow), specific signals for retention have been described for some ER and Golgi proteins (Figure 5, Table 1). Perhaps the best defined retention signal is the tetrapeptide His/Lys-Asp-Glu-Leu (HDEL or KDEL), which is present at the C-terminus of several soluble resident ER proteins, in animals, yeast, and plants (see later discussion).

Sorting information is required to divert proteins from bulk-flow transport to the yeast cell vacuole (via sorting in the TGN); recent evidence suggests a similar requirement for transport of proteins to the plant cell vacuole. As mentioned earlier, a chimeric gene consisting of the signal sequence of the vacuolar seed protein PHA and the coding sequence of a cytosolic seed protein (pea albumin 2) has been expressed in seeds and suspension cultured cells of transgenic tobacco.[62,63] The signal peptide is necessary and sufficient to direct entry of the chimeric protein into the secretory pathway; however, it does not accumulate in the vacuole (e.g., in seeds) and is efficiently secreted from suspension cultured cells.

In animal cells, proteins *en route* to the lysosome share a common pathway of intracellular

transport with plasma membrane and constitutively secreted proteins that extends as far as the TGN (reviewed in Reference 85). However, the fate of soluble lysosomal proteins (e.g., lysosomal hydrolases) that are all glycosylated in the ER, is determined earlier in the Golgi stack. After arrival in the *cis*-Golgi, a series of modifications by specific Golgi enzymes (*N*-acetylglucosamine phosphotransferase and phosphoglycosidase) ultimately result in a mannose-6-phosphate tag on soluble lysosomal proteins. The specificity of the reaction lies with the enzyme *N*-acetylglucosamine phosphotransferase, which recognizes a conformation-dependent signal patch on lysosomal hydrolases, such that only proteins destined to reside in lysosomes are appropriately modified (reviewed in Reference 6). Later in the pathway, the mannose-6-phosphate tag allows the proteins to bind to mannose-6-phosphate receptors and hence be specifically sorted for delivery to the lysosome in the TGN. Release of lysosomal proteins from their receptors occurs in an acidified (pre-lysosomal, endosomal) compartment; subsequently, the receptors recycle back to the Golgi (or to the cell surface because these receptors are also involved in endocytosis) while the proteins are packaged into lysosomes.

In animal cells, there are two distinct pathways of secretion viz., constitutive and regulated.[16,56,86,87] In specialized cells involved in regulated secretion (that are also capable of secreting proteins constitutively), the regulated secretory products are concentrated and stored in dense vesicles (granules) in the cytosol. Fusion of these vesicles with the cell surface occurs only after receipt of a specific extracellular signal (e.g., a hormone or neurotransmitter, depending on the cell type). This process, termed exocytosis,[88] is thus regulated, and gives rise to the controlled secretion of the selected contents of the secretory storage vesicle.[56,57] By contrast, secretion via bulk flow to the cell surface is constitutive and continuous. Characterization of the signals that permit membership or entry of specific proteins into storage vesicles is not yet complete; however, it is clear that their content is highly selected. A form of selective precipitation in the TGN may be involved; "carrier-type" proteins have been implicated in the mechanism.[89]

In plants, the constitutive pathway of protein secretion is the principal pathway along which proteins are secreted to the cell exterior;[17,29] in general, there is a close temporal correlation between the synthesis of proteins and their secretion from the cell (e.g., cereal α-amylase).[90,91]

Regulated secretion occurs in the green alga *Clamydomonas reinhardii*. Here the release of stored lysin (a cell-wall-degrading enzyme) is induced by gamete sexual signaling.[92]

As pointed out by Satiat-Jeunemaitre and Hawes,[3] the terms 'regulated' and 'constitutive' are somewhat misleading; constitutive secretion is regulated in the sense that it is influenced by a number of external factors such as hormones and environmental stresses (see later discussion).

a. Signal-Dependent Translocation Across the ER Membrane

Translocation across the membrane of the rough ER represents the first signal-dependent step along the secretory pathway. Proteins destined for this step contain a hydrophobic signal peptide (Figure 1, Table 1) that serves as a requisite signal to target the ribosome to the rough ER. The development of *in vitro* protein translation-translocation systems has been important for the characterization of some of the molecular components that regulate ER targeting/translocation in eukaryotes (reviewed in References 10 and 93). Cell-free translation systems have been prepared from lysates of wheat-germ, reticulocytes, and yeast. They contain the ribosomes and additional co-factors required for protein biosynthesis, but need to be supplemented with an energy source; typically, a source of exogenous amino acids, including a radiolabelled amino acid such as [^{35}S]-methionine is added along with mRNAs. Reaction mixtures can be supplemented with vesiculated portions of the ER (microsomes) that are isolated by homogenizing cells in isotonic buffers, then purified from other membrane components by centrifugation on sucrose density gradients (Figure 6).[94] In this way, the nascent polypeptide chain is targeted and inserted into the lumen of the microsomes where

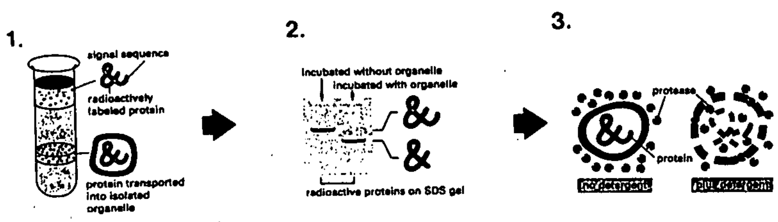


FIGURE 6. Biochemical approaches for studying the mechanism of protein translocation. In this approach a labeled protein containing a specific signal sequence is transported into isolated organelles *in vitro*. The labeled protein is usually produced by cell-free translation of a purified mRNA encoding the protein; radioactive amino acids are used to label the newly synthesized protein so that it can be distinguished from the many other proteins that are present in the *in vitro* translation system. Three methods are commonly used to test if the labeled protein has been translocated into the organelle: (1) The labeled protein co-fractionates with the organelle during centrifugation; (2) The signal sequence is removed by a specific protease that is present inside the organelle; and (3) The protein is protected from digestion when proteases are added to the incubation medium, but is susceptible if a detergent is first added to disrupt the organelle membrane. By exploiting such *in vitro* assays, one can determine what components (proteins, ATP, GTP, etc.) are required for the translocation process. (From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D., *Molecular Biology of the Cell*, 3rd ed., Garland Publishing, New York, 1994. With permission from Garland Publishing.)

the signal peptide is cleaved off, producing a decrease in the molecular mass of the protein. The products of translation and the effects of translocation can be viewed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography following an immunopurification step (Figure 6). Translocation is demonstrated by the addition of proteases; proteins within the microsomes are protected from digestion unless solubilized with detergent.

Experiments to establish the autonomous nature of signal peptides and their role in ER targeting/translocation typically involve the generation of a chimeric gene construct (to encode a chimeric mRNA and resultant protein), in which DNA encoding a signal peptide is linked to that encoding a passenger (e.g., cytosolic) protein. This chimeric sequence can then be ligated into a plasmid downstream of an active promoter that is recognized by SP6 RNA polymerase (Figure 7). When the resultant mRNA produced by this *in vitro* transcription system is translated in the cell-free system supplemented with microsomes, the protein is translocated across the microsome (ER) membrane as confirmed by protease treatment (Figures 6 and 7).

In vitro protein import systems (as well as subcellular localization of chimeric proteins synthesized in stably transformed tobacco; see earlier) have established that the signal peptides of plant proteins generally function with a high degree of autonomy, being both necessary and sufficient to direct passenger proteins into the lumen of the ER. An examination of the derived amino acid sequences of plant secretory and vacuolar proteins shows that all of them have a N-terminal domain with the properties of other eukaryotic signal sequences.[17,95–97] Despite great sequence diversity between eukaryotic signal peptides, most have similar overall features viz., a charged N-terminal region, a core of 10 to 15 hydrophobic amino acids, and a more polar C-terminal sequence (Figure 8).[98] Translocation is mediated by receptor-like systems[54,99,100] (termed the translocon) that involve proteins in the cytosol and on the target membrane; this complex cellular machinery is able to recognize signal peptides, despite their variation in amino acid sequence. Studies on the well-characterized mammalian ER have identified several distinct components of such a receptor-like system; these act sequentially and, hence, greatly enhance the specificity of the sys-

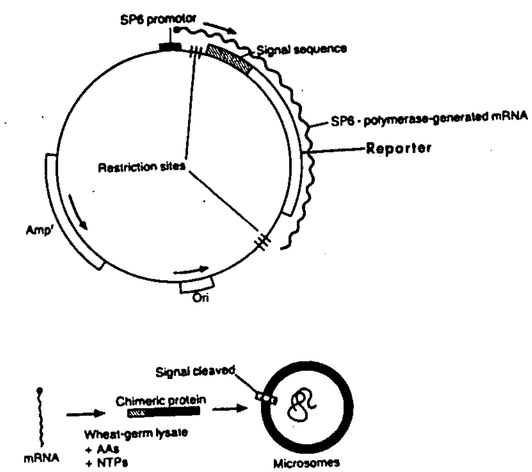


FIGURE 7. Generation of a chimeric protein. DNA encoding a signal sequence ligated to DNA encoding a normally cytoplasmic protein is transcribed and translated to yield a precursor protein that is translocated into microsomal vesicles. The signal sequence is cleaved off on translocation. (From Austen, B. M. and Westwood, O. M. R., *Protein Targeting and Secretion*, Oxford University Press, Oxford, 1991. With permission from Oxford University Press.)

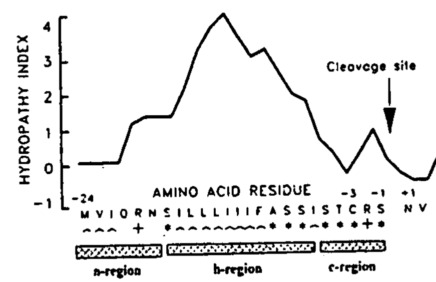


FIGURE 8. Structure and hydropathy index of the signal peptide of the plant protein, tomato fruit polygalacturonase. The sequence was deduced from the nucleotide sequence of the cDNA and the signal sequence cleavage site by radiosequencing of the product of *in vitro* translation and *in vitro* processing (see Bennett and Osteryoung, 1991 and references therein). The n-region is characterized by a single positively charged amino acid (+), the h-region by a core of hydrophobic amino acids (^), and the c-region by small neutral amino acids (*) at positions -1 and -3. (From Bennett, A.B., and Osteryoung, K.W., in Plant Genetic Engineering, Grierson, D., Ed., Chapman and Hall, New York, 1991, 199. With permission from Chapman and Hall.)

tem (reviewed in References 5, 15, 93, 101 and 102). One component is the signal recognition particle (SRP), a ribonucleoprotein complex consisting of a molecule of RNA and six polypeptide chains, including a 54-kDa protein with a guanosine trisphosphate (GTP)-binding domain. The 54-kDa polypeptide subunit (SRP-54) selectively binds to the signal peptide sequence on the precursor as it emerges from the ribosome, temporarily halting translation. The ternary complex (consisting of the ribosome, nascent precursor, and SRP) is targeted to the rough ER membrane due to the affinity of the SRP for a 72-kDa integral ER protein (part of the SRP receptor or docking protein). Contact between the SRP and its receptor results in the dissociation of SRP from both the ribosome and the signal sequence. The ribosome becomes membrane bound (possibly by associating with several ER proteins), translation resumes, and the signal sequence is inserted into the translocation site to begin the translocation event. Subsequently, the SRP is released from its receptor (by hydrolysis of a bound GTP molecule), and the SRP is free to enter a new targeting cycle (Figure 9). At least two steps of the translocation event appear to be controlled by the hydrolysis of GTP, again contributing to the specificity of the system.[101] The GTPase cycle of SRP-54 may control signal sequence insertion into the translocation channel.[102] The ER-translocation process in plants is similar to that in animals and fungi. The existence of the SRP and SRP receptor in plants has now been documented. Moreover, efficient synthesis, translocation, and processing of plant secretory proteins occur in animal and fungal heterologous systems; thus, strong evidence exists for a common mechanism of protein translocation in eukaryotes.[91,103,104] The 54-kDa protein subunit of the SRP of *Arabidopsis* is encoded by at least two genes and exhibits a high degree of sequence similarity to the SRP-54 polypeptides of mammals and yeast.[105] The functional domains characteristic of the yeast and mammalian proteins are also present, viz., a N-terminal GTP-binding (G-) domain and a C-terminal methionine-rich (M-) domain, capable of forming amphiphilic α-helices (see below).

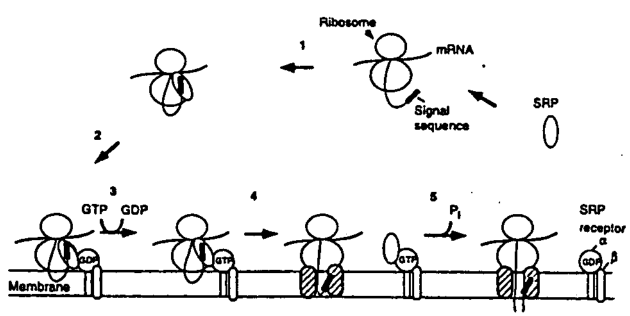


FIGURE 9. The ER targeting cycle. The scheme shows the first steps in protein translocation across the ER membrane. When the signal sequence of a growing polypeptide chain has emerged from the ribosome, it is recognized and bound by the SRP (step 1). In step 2, the complex containing the ribosome, nascent chain, and SRP binds to the ER membrane through an interaction of the SRP with its membrane receptor, which consists of two subunits. In step 3, GDP is exchanged for GTP at the SRP receptor. In step 4, the SRP is released from the ribosome and from the signal sequence. The ribosome becomes membrane-bound, probably by its interaction with SEC61 or associated proteins, and translocation begins. In step 5, the SRP is released from its receptor by GTP hydrolysis and can then enter a new targeting cycle. Whether the two subunits of the SRP receptor dissociate during the SRP cycle is not known. (From Rapoport, T. A., *Science*, 258, 931, 1992. With permission from T. A. Rapoport.)

Solutions to the enigma of how diverse signal peptide sequences/shapes can be recognized by, and accommodated within, the 54-kDa polypeptide component of the SRP have been sought (reviewed in Reference 101). Sequestration of signal peptide sequences may occur in a hydrophobic methionine-rich pocket formed by secondary structural domains (amphipathic α-helices) of the 54-kDa SRP protein (the M-domain). More specifically, the unbranched methionine side chains lining this pocket provide projecting flexible "bristles" that recognize and accommodate only the common hydrophobic core of signal peptides, albeit diverse in primary sequence.[93,101,102,106] Less is known about the mechanics of protein translocation; transport through a proteinaceous pore (rather than movement directly through the lipid bilayer) is suggested by studies in yeast and mammals.[107,108] Electrophysiological techniques indicate protein conducting channels of discrete size in rough ER-derived vesicles.[109] Likewise, analyses using proteins tagged with fluorescent amino acid analogs indicate that the nascent polypeptide resides in an aqueous environment following its insertion into the translocation site.[110] A tight seal between the ribosome and the translocation site restricts solute flow across the membrane; both these properties are consistent with a hydrophilic, proteinaceous translocation site. The growing number of putative translocon components identified in various eukaryotic organisms (Table 3) attests to the complexity of the translocation machinery. Some of these components are directly involved in the translocation process per se; others likely participate in chemical modifications of a nascent polypeptide or in its folding and assembly. In yeast, identification of proteins comprising the central core of the translocon came from genetic selection to obtain yeast mutants defective in translocation and causing the cytoplasmic accumulation of precursors for secretory proteins *in vivo* and *in vitro*. In this way, SEC61, SEC62, and SEC63 were identified as integral membrane proteins assembled together with two additional polypeptides into a multisubunit complex in the ER.[111,112] Crosslinking studies (see below) indicate a direct role in translocation for SEC61; this protein is cross-linked to polypeptides trapped in translocation.[113] Mutations in *sec62* and *sec63* genes decrease the ability of the SEC61 protein to interact with translocating polypeptides, and a role during the early stages of the translocation process has been proposed. Another protein called the binding protein (BiP) (a luminal ER protein; see later discussion) is thought by some to play a facilitory role in the translocation of precursors in yeast. More specifically, protein entry into the ER lumen may be facilitated by cycles of BiP binding and adenosine triphosphate (ATP) hydrolysis-dependent dissociation.[102] The protein may be relatively more important for the posttranslational mode of translocation (e.g., in yeast), determining the directionality of transport by binding to the incoming protein and preventing the inappropriate release/exit of proteins from the cytoplasmic face of the translocation channel. A role in the folding of translocation components

TABLE 3
Possible Components of the Translocation Site of the ER Membrane[a]

Protein	Occurrence	Function
SEC61/Y	Yeast, bacteria, mammals, fish	Constituent of a protein-conducting channel
TRAM protein	Mammals	Early function in translocation
SEC62-SEC63 complex	Yeast	Early function in translocation
SSR complex	Mammals, fish, birds	Unknown
Signal peptidase complex	Yeast, mammals	Signal peptide cleavage
Oligosaccharyl transferase	Mammals, yeast	Asn-glycosylation
mp30	Mammals	Unknown

[a] From Rapoport, T. A., *Science*, 258, 931, 1992. With permission from T. A. Rapoport. See references therein.

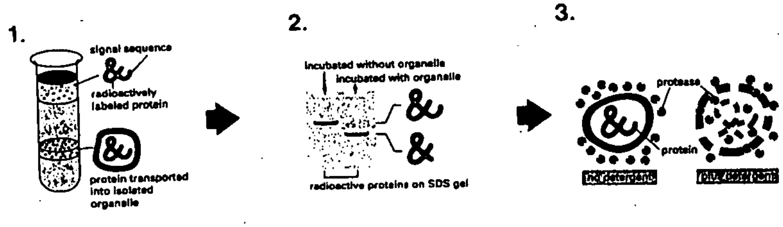


FIGURE 6. Biochemical approaches for studying the mechanism of protein translocation. In this approach a labeled protein containing a specific signal sequence is transported into isolated organelles *in vitro*. The labeled protein is usually produced by cell-free translation of a purified mRNA encoding the protein; radioactive amino acids are used to label the newly synthesized protein so that it can be distinguished from the many other proteins that are present in the *in vitro* translation system. Three methods are commonly used to test if the labeled protein has been translocated into the organelle: (1) The labeled protein co-fractionates with the organelle during centrifugation; (2) The signal sequence is removed by a specific protease that is present inside the organelle; and (3) The protein is protected from digestion when proteases are added to the incubation medium, but is susceptible if a detergent is first added to disrupt the organelle membrane. By exploiting such *in vitro* assays, one can determine what components (proteins, ATP, GTP, etc.) are required for the translocation process. (From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D., *Molecular Biology of the Cell*, 3rd ed., Garland Publishing, New York, 1994. With permission from Garland Publishing.)

the signal peptide is cleaved off, producing a decrease in the molecular mass of the protein. The products of translation and the effects of translocation can be viewed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography following an immunopurification step (Figure 6). Translocation is demonstrated by the addition of proteases; proteins within the microsomes are protected from digestion unless solubilized with detergent.

Experiments to establish the autonomous nature of signal peptides and their role in ER targeting/translocation typically involve the generation of a chimeric gene construct (to encode a chimeric mRNA and resultant protein), in which DNA encoding a signal peptide is linked to that encoding a passenger (e.g., cytosolic) protein. This chimeric sequence can then be ligated into a plasmid downstream of an active promoter that is recognized by SP6 RNA polymerase (Figure 7). When the resultant mRNA produced by this *in vitro* transcription system is translated in the cell-free system supplemented with microsomes, the protein is translocated across the microsome (ER) membrane as confirmed by protease treatment (Figures 6 and 7).

In vitro protein import systems (as well as subcellular localization of chimeric proteins synthesized in stably transformed tobacco; see earlier) have established that the signal peptides of plant proteins generally function with a high degree of autonomy, being both necessary and sufficient to direct passenger proteins into the lumen of the ER. An examination of the derived amino acid sequences of plant secretory and vacuolar proteins shows that all of them have a N-terminal domain with the properties of other eukaryotic signal sequences.[17,95–97] Despite great sequence diversity between eukaryotic signal peptides, most have similar overall features viz., a charged N-terminal region, a core of 10 to 15 hydrophobic amino acids, and a more polar C-terminal sequence (Figure 8).[98] Translocation is mediated by receptor-like systems[54,99,100] (termed the translocon) that involve proteins in the cytosol and on the target membrane; this complex cellular machinery is able to recognize signal peptides, despite their variation in amino acid sequence. Studies on the well-characterized mammalian ER have identified several distinct components of such a receptor-like system; these act sequentially and, hence, greatly enhance the specificity of the sys-

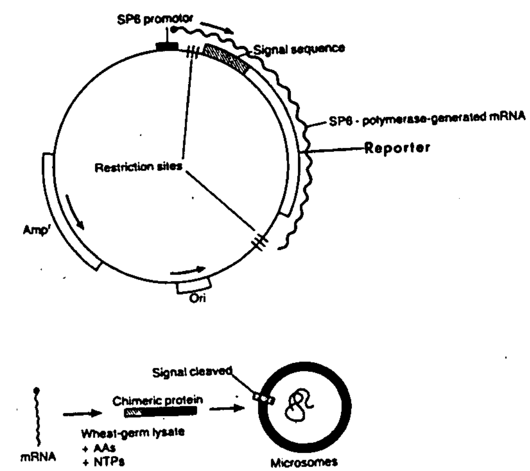


FIGURE 7. Generation of a chimeric protein. DNA encoding a signal sequence ligated to DNA encoding a normally cytoplasmic protein is transcribed and translated to yield a precursor protein that is translocated into microsomal vesicles. The signal sequence is cleaved off on translocation. (From Austen, B. M. and Westwood, O. M. R., *Protein Targeting and Secretion*, Oxford University Press, Oxford, 1991. With permission from Oxford University Press.)

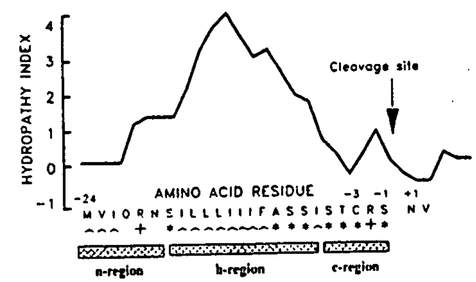


FIGURE 8. Structure and hydropathy index of the signal peptide of the plant protein, tomato fruit polygalacturonase. The sequence was deduced from the nucleotide sequence of the cDNA and the signal sequence cleavage site by radiosequencing of the product of *in vitro* translation and *in vitro* processing (see Bennett and Osteryoung, 1991 and references therein). The n-region is characterized by a single positively charged amino acid (+), the h-region by a core of hydrophobic amino acids (^), and the c-region by small neutral amino acids (*) at positions -1 and -3. (From Bennett, A.B., and Osteryoung, K.W., in Plant Genetic Engineering, Grierson, D., Ed., Chapman and Hall, New York, 1991, 199. With permission from Chapman and Hall.)

tem (reviewed in References 5, 15, 93, 101 and 102). One component is the signal recognition particle (SRP), a ribonucleoprotein complex consisting of a molecule of RNA and six polypeptide chains, including a 54-kDa protein with a guanosine trisphosphate (GTP)-binding domain. The 54-kDa polypeptide subunit (SRP-54) selectively binds to the signal peptide sequence on the precursor as it emerges from the ribosome, temporarily halting translation. The ternary complex (consisting of the ribosome, nascent precursor, and SRP) is targeted to the rough ER membrane due to the affinity of the SRP for a 72-kDa integral ER protein (part of the SRP receptor or docking protein). Contact between the SRP and its receptor results in the dissociation of SRP from both the ribosome and the signal sequence. The ribosome becomes membrane bound (possibly by associating with several ER proteins), translation resumes, and the signal sequence is inserted into the translocation site to begin the translocation event. Subsequently, the SRP is released from its receptor (by hydrolysis of a bound GTP molecule), and the SRP is free to enter a new targeting cycle (Figure 9). At least two steps of the translocation event appear to be controlled by the hydrolysis of GTP, again contributing to the specificity of the system.[101] The GTPase cycle of SRP-54 may control signal sequence insertion into the translocation channel.[102] The ER-translocation process in plants is similar to that in animals and fungi. The existence of the SRP and SRP receptor in plants has now been documented. Moreover, efficient synthesis, translocation, and processing of plant secretory proteins occur in animal and fungal heterologous systems; thus, strong evidence exists for a common mechanism of protein translocation in eukaryotes.[91,103,104] The 54-kDa protein subunit of the SRP of *Arabidopsis* is encoded by at least two genes and exhibits a high degree of sequence similarity to the SRP-54 polypeptides of mammals and yeast.[105] The functional domains characteristic of the yeast and mammalian proteins are also present, viz., a

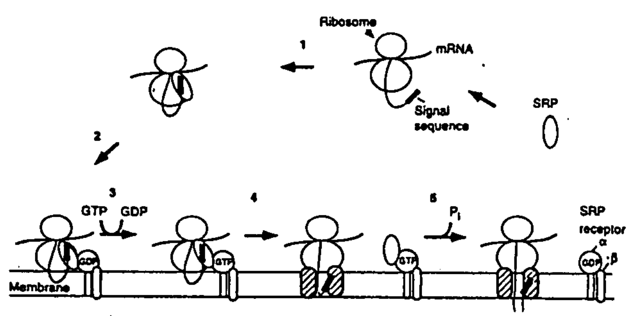


FIGURE 9. The ER targeting cycle. The scheme shows the first steps in protein translocation across the ER membrane. When the signal sequence of a growing polypeptide chain has emerged from the ribosome, it is recognized and bound by the SRP (step 1). In step 2, the complex containing the ribosome, nascent chain, and SRP binds to the ER membrane through an interaction of the SRP with its membrane receptor, which consists of two subunits. In step 3, GDP is exchanged for GTP at the SRP receptor. In step 4, the SRP is released from the ribosome and from the signal sequence. The ribosome becomes membrane-bound, probably by its interaction with SEC61 or associated proteins, and translocation begins. In step 5, the SRP is released from its receptor by GTP hydrolysis and can then enter a new targeting cycle. Whether the two subunits of the SRP receptor dissociate during the SRP cycle is not known. (From Rapoport, T. A., *Science*, 258, 931, 1992. With permission from T. A. Rapoport.)

N-terminal GTP-binding (G-) domain and a C-terminal methionine-rich (M-) domain, capable of forming amphiphilic α-helices (see below).

Solutions to the enigma of how diverse signal peptide sequences/shapes can be recognized by, and accommodated within, the 54-kDa polypeptide component of the SRP have been sought (reviewed in Reference 101). Sequestration of signal peptide sequences may occur in a hydrophobic methionine-rich pocket formed by secondary structural domains (amphipathic α-helices) of the 54-kDa SRP protein (the M-domain). More specifically, the unbranched methionine side chains lining this pocket provide projecting flexible "bristles" that recognize and accommodate only the common hydrophobic core of signal peptides, albeit diverse in primary sequence.[93,101,102,106] Less is known about the mechanics of protein translocation; transport through a proteinaceous pore (rather than movement directly through the lipid bilayer) is suggested by studies in yeast and mammals.[107,108] Electrophysiological techniques indicate protein conducting channels of discrete size in rough ER-derived vesicles.[109] Likewise, analyses using proteins tagged with fluorescent amino acid analogs indicate that the nascent polypeptide resides in an aqueous environment following its insertion into the translocation site.[110] A tight seal between the ribosome and the translocation site restricts solute flow across the membrane; both these properties are consistent with a hydrophilic, proteinaceous translocation site. The growing number of putative translocon components identified in various eukaryotic organisms (Table 3) attests to the complexity of the translocation machinery. Some of these components are directly involved in the translocation process per se; others likely participate in chemical modifications of a nascent polypeptide or in its folding and assembly. In yeast, identification of proteins comprising the central core of the translocon came from genetic selection to obtain yeast mutants defective in translocation and causing the cytoplasmic accumulation of precursors for secretory proteins *in vivo* and *in vitro*. In this way, SEC61, SEC62, and SEC63 were identified as integral membrane proteins assembled together with two additional polypeptides into a multisubunit complex in the ER.[111,112] Crosslinking studies (see below) indicate a direct role in translocation for SEC61; this protein is cross-linked to polypeptides trapped in translocation.[113] Mutations in *sec62* and *sec63* genes decrease the ability of the SEC61 protein to interact with translocating polypeptides, and a role during the early stages of the translocation process has been proposed. Another protein called the binding protein (BiP) (a luminal ER protein; see later discussion) is thought by some to play a facilitory role in the translocation of precursors in yeast. More specifically, protein entry into the ER lumen may be facilitated by cycles of BiP binding and adenosine triphosphate (ATP) hydrolysis-dependent dissociation.[102] The protein may be relatively more important for the posttranslational mode of translocation (e.g., in yeast), determining the directionality of transport by binding to the incoming protein and preventing the inappropriate release/exit of proteins from the cytoplasmic face of the translocation channel. A role in the folding of translocation components

TABLE 3
Possible Components of the Translocation Site of the ER Membrane[a]

Protein	Occurrence	Function
SEC61/Y	Yeast, bacteria, mammals, fish	Constituent of a protein-conducting channel
TRAM protein	Mammals	Early function in translocation
SEC62-SEC63 complex	Yeast	Early function in translocation
SSR complex	Mammals, fish, birds	Unknown
Signal peptidase complex	Yeast, mammals	Signal peptide cleavage
Oligosaccharyl transferase	Mammals, yeast	Asn-glycosylation
mp30	Mammals	Unknown

[a] From Rapoport, T. A., *Science*, 258, 931, 1992. With permission from T. A. Rapoport. See references therein.

like the SEC62-SEC63 protein complex has also been suggested.[93]

Identification of potential components of the mammalian translocon has also been aided by the use of photochemicals or chemicals that cross-link nascent secretory proteins to adjacent integral ER membrane proteins. These studies implicated a 35- to 39-kDa glycoprotein (translocating chain-associated membrane [TRAM] protein) and a 34- to 37-kDa nonglycoprotein (homologous to yeast SEC61 and *Escherichia coli* SECY proteins).[114,115]

Reconstitution of detergent-solubilized rough ER proteins into translocation competent proteoliposomes allows the functional evaluation of putative translocation factors.[108,114] In mammalian cells, the SRP, the SRP receptor, SEC61, and the TRAM protein may participate directly in translocation because translocation is successful using these purified components. However, several additional stimulatory factors probably exist (perhaps mammalian homologs of the yeast proteins, SEC62, and SEC63). The TRAM protein may contact the N-terminal regions of nascent polypeptides at an early stage; a requirement for this protein depends upon the synthesized product. For example, it is essential for the transport of several precursors, but is only stimulatory for others. Further components assumed to be involved in the process include putative ribosome receptors of 34 kDa[116,117] and 180 kDa,[118] a membrane protein of 30 kDa with affinity for SRP (mp30),[119] and an unidentified ATP-binding membrane protein.[120] However, the function of these components, if any, is either disputed or not demonstrated yet (Figure 10) (see Reference 108 and references therein).

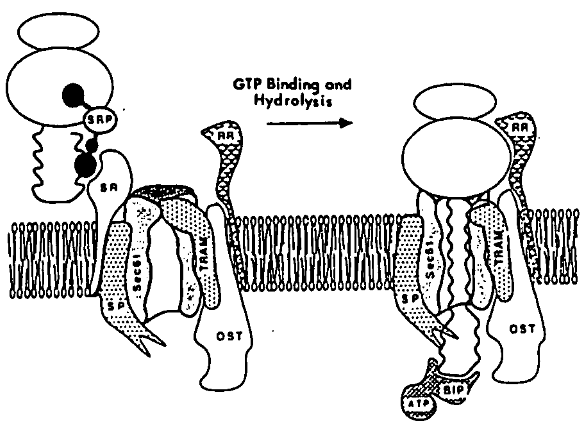


FIGURE 10. A schematic representation of the targeting and transport phases of a protein translocation reaction. The left portion of the diagram depicts SRP-dependent targeting of a ribosome to a preassembled translocation site comprised of the SRP receptor (SR), SEC61, TRAM, signal peptidase (SP), oligosaccharyltransferase (OST), and a ribosome receptor (RR). Conformational changes in Sec61p and TRAM that mediate channel gating are speculative. GTP binding and hydrolysis reactions permit SRP dissociation from the signal sequence and SRP receptor. Polypeptides traversing the channel contact TRAM and SEC61. Luminal reactions mediated by signal peptidase and oligosaccharyltransferase may inhibit reversible transport through the channel, whereas BiP may actively facilitate transport by binding the precursor as it emerges into the lumen. (From Gilmore, R., *Cell*, 75, 589, 1993. With permission from Cell Press.)

A hydrophilic protein is completely translocated across the rough ER if it is preceded by a signal that undergoes cleavage as translocation is occurring. However, in cases where the signal peptide is located internally, or its hydrophobic core is integrated within a longer membrane anchor domain, cleavage does not occur[15] (see later discussion). In this latter case, the "stop-transfer" sequence halts translocation and remains embedded in the membrane, functioning as the anchor for integral membrane proteins. It has been suggested that the proteinaceous ER translocation channel contains a receptor for stop-transfer sequences, the channel opening laterally to allow integration of membrane proteins.[93,102]

Protein synthesis (i.e., translation) and translocation of the protein generally occur simultaneously. Strong evidence argues against a mechanistic coupling between these two events; however, their temporal association may nonetheless play an important role — that of ensuring that nascent polypeptides maintain (or acquire) a translocation-competent conformation.[5] In mammalian cells, binding of the SRP to the signal peptide and the arrest of translation until functional insertion into the ER occurs would largely perform this important function,[15] although accessory proteins have also been implicated. SRP-like proteins have been detected in yeast. But this organism may also have a SRP-independent targeting pathway because it can survive without the SRP, and various proteins can be translocated *in vitro* in the absence of the SRP and the SRP receptor.[121] Any posttranslational translocation must occur without the SRP because the latter binds only to nascent chains attached to ribosomes.[122] It is likely that cytosolic chaperones such as 70-kDa heat-shock proteins (and probably additional proteins) function to assist proteins in maintaining a translocation-competent state.[107,123,124]

In most cases, the signal peptide is removed (in a cotranslational manner) from the nascent polypeptide chain as it emerges into the ER lumen by a signal peptidase located on the inner face of the membrane. Cleavage generally occurs at the more polar C-terminal region of the signal sequence and is thought to allow subsequent processes like protein folding and assembly to proceed normally (reviewed in Reference 4). For example, when the soybean storage protein glycinin is synthesized in an *in vitro* translation system (in the absence of ER membranes) from a synthetic mRNA corresponding to the entire glycinin coding region, the presence of the signal peptide inhibits assembly (trimerization) and results in misfolding of monomers.[125] The signal peptide has a similar negative effect on the acquisition of a functional (biologically active) conformation of α-amylase.[126]

b. Signal-Dependent Retention of Soluble Proteins in the ER Lumen

As mentioned previously, soluble resident proteins of the ER are distinguished by a carboxy-terminal tetrapeptide sequence. In mammals, these sequences are typically K/H/RDEL[59,70,127] and H/DDEL in yeasts.[127,128] Some of the more abundant soluble resident ER proteins of eukaryotes identified with these carboxy-terminal sequences include BiP, protein disulfide isomerase (PDI), and others.[129–131] In plants, the sequences have been identified at the carboxy-termini of several ER proteins (e.g., BiP, in tomato, maize, and tobacco; an auxin-binding protein in maize; and PDI in alfalfa).[17,59,132–142] The carboxy-terminal tetrapeptide sequence is necessary for retention of proteins in the ER.[70,143,144] When the sequence is deleted from an ER-resident protein (e.g., BiP), secretion is the outcome;[145] conversely, addition of the tetrapeptide to the carboxy-terminus of lysozyme (a protein that is normally secreted) or cathepsin D (a lysosomal enzyme) causes the proteins to accumulate in the ER.[70,143] Although some changes to individual amino acids can be tolerated, extending the sequence with random amino acids at the carboxy-terminus (or changing KDEL to KDAS) results in secretion of the target protein.[70,146] Thus, retention may be mediated by a specific receptor that recognizes (H/K/R)DEL only when it is present at the extreme carboxy-terminus of a protein (but see later discussion). In addition to KDEL, the carboxy-terminal sequences HDEL and RDEL are sufficient for retention in the ER and nuclear envelope of the reporter enzyme phosphinothricin acetyl transferase (PAT) in stably and transiently transformed plant cells

from a variety of species,[140,141] indicating that the plant ER retention mechanism can recognize several sequences. However, changing certain sequences within KDEL (such as KEEL, SDEL, KDDL, KDEI, and KDEV), leads to partial or complete loss of ER retention.

The nature of ER retention signals and the machinery involved in their recognition may be more complex than originally anticipated and may be distinct (in certain aspects) in different eukaryotic organisms. In addition to KDEL, HDEL, and RDEL, a variety of other sequences occur at the C-termini of ER-resident proteins in plants, yeast and mammals. Likewise, several other sequences are able to relocalize passenger proteins to the ER lumen in these eukaryotic organisms (Table 4) (reviewed in Reference 4; see references therein). Moreover, such unrelated sequences as the conserved *s*-cyclophilin-specific C-terminal region VEKPFAIAKE is sufficient to direct a secretory protein toward ER subcompartments.[147] In addition, there are examples of proteins that contain a carboxyterminal KDEL sequence (as deduced from their cDNA sequences), but are found in sites other than the ER lumen. An example is the plant auxin binding protein (ABP1) that is secreted (via the Golgi complex) to the cell surface (the plasma membrane and cell wall).[148] Interaction between the ER retention receptor and the KDEL tetrapeptide on this protein is suggested to be conditional. It may be regulated by auxin because binding of this hormone appears to change the structure of the protein (possibly its C-terminus and/or the accessibility of the KDEL retention signal), as revealed by an altered affinity of the auxin-ABP1 complex to monoclonal anti-ABP1-KDEL antibodies.[149] However, more information is needed to confirm these possibilities. PDI, containing a KDEL sequence at its C-terminal end, is not only found within the ER lumen but is also transported to the plasma membrane via the secretory pathway in rat exocrine pancreatic cells.[150] Thus, it is becoming increasingly apparent that the nature of ER localization signals is not obvious from the primary protein structure of ER-resident soluble proteins. Other approaches such as analyses of the crystal structures of R/H/KDEL-containing proteins[151,152] and nuclear magnetic resonance (NMR) analyses of a variety of peptides comprising the C-termini of ER resident proteins may be necessary to elucidate the critical features of ER localization signals.[4]

TABLE 4
Tetrapeptides Occurring in ER Proteins (A) and Comparative Analysis of Sequence Motifs Using Carrier Proteins (B)

	Mammalian cells	Plant cells
(A)	KDEL	KDEL
	RDEL	HDEL
	HDEL	NDELK
	KEEL	KQEL
	QEDL	
	HTEL	
	HIEL	
(B)	KDEL	KDEL
	RDEL	RDEL
	HDEL	HDEL
	KEEL	KEEL
	KEDL	SDEL
	KDDL	
	KNEL	
	REEL	
	REDL	
	RDDL	
	QDEL	
	QEEL	
	DKEL	
	KDEI	

From Vitale, A., Ceriotti, A., and Denecke, J., *J. Exp. Bot.*, 44, 1417, 1993. With permission from Oxford University Press. See references therein.

The addition of KDEL or HDEL onto the carboxy-terminus of some proteins retards transport, but is not sufficient for their absolute retention in the ER.[153,154] Retardation of export rather than total ER retention occurs with several mammalian secretory proteins that are modified at the exact carboxy-terminus to encode the sequence SEKDEL.[153] When the carboxy-terminal sequence of a plant vacuolar protein (PHA) is changed from LNQIL to LNKDEL and the derivative protein expressed in transgenic tobacco, a large proportion is localized in the ER and the nuclear envelope (that is continuous with the ER). However, a considerable amount of the PHA-KDEL is transported through the Golgi complex and reaches the protein storage vacuole/protein body.[154] Partial retention of these proteins in the ER may be due to a less than optimal display of the carboxy-terminal KDEL sequence, thus diminishing its recognition by the appropriate receptor[155] and allowing its escape from the normal retrieval mechanism (see subsequent discussion). Other structural features of the protein may be required for ER retention, and their role may not be restricted to simply ensuring optimal positioning of the tetrapeptide for receptor-mediated recognition.[4,17]

Changing the carboxy-terminus of the vacuolar pea storage protein vicilin to include KDEL (SEKDEL) results in a dramatic increase in the accumulation of vicilin in the leaves of transgenic tobacco and alfalfa.[156,157] The retention signal keeps the protein from advancing to a compartment where it could be degraded (presumably the vacuole) and results in the formation of protein-body-like inclusions. These ER-derived, membrane-bound, electron-dense structures (0.5 to 1.0 µm in diameter) contain vicilin and resemble the ER-derived protein bodies found in the endosperm cells of certain cereals such as maize and sorghum (see Section V.G.1).

Given that most proteins are transported by bulk flow, whereas certain proteins are selectively retained, it is pertinent to examine the mechanism of specific retention in the ER. The luminal proteins are not associated with a membrane-bound receptor,[59,145] nor is retention guaranteed by membrane anchoring; many membrane-bound viral proteins are transported out of the ER.[60] It appears that for soluble proteins at least, ER retention mechanisms come into play at a later point in the secretory pathway, that is, after the proteins have been transported out of the ER.[59,143] More specifically, luminal ER proteins that escape are continuously and specifically retrieved from a post-ER ("salvage") compartment (that is pre-Golgi or within the Golgi complex) by a receptor;[158] the receptor-ligand complexes are cycled back to the ER (by specialized vesicular transport; see subsequent discussion), where the KDEL-bearing proteins are released.[59] In yeast (*Saccharomyces cerevisiae*), the retrieval mechanism appears to be in the Golgi complex,[127,159] and some very elegant studies in this organism[160-162] have identified a putative receptor involved in HDEL recognition. A genetic approach was utilized to identify the essential components of the "recycling" pathway; to this end, *S. cerevisiae* mutants specifically defective in the HDEL-mediated ER retention system were isolated, allowing subsequent identification of two genes required in the process, viz., *erd1* and *erd2* (for "ER retention defective"). Deletion of the *erd1* gene causes a pleiotropic defect in part of the Golgi apparatus, which evidently results in inefficient retrieval from this organelle.[161] The *erd2* gene encodes a 26-kDa integral membrane protein that is required both for retention of ER proteins and perhaps indirectly for normal traffic of proteins through the Golgi.[160] Strikingly, the abundance of this protein determines both the efficiency and capacity of the retention system: reduced expression leads to secretion of HDEL-tagged proteins; conversely, its overexpression improves their retention (both in wild-type cells and in other mutants). Moreover, *erd2* determines the signal specificity of the retention system;[162] exchange of the *erd2* gene from *S. cerevisiae* (which only recognizes HDEL) for the corresponding gene from another yeast, *Kluyveromyces lactis* (which recognizes either HDEL or DDEL), allows equal recognition of DDEL and HDEL in *S. cerevisiae*. These observations, as well as the subcellular localization of the protein (i.e., in a post-ER, Golgi-like compartment), make it an excellent candidate for the receptor that sorts lumenal ER proteins in yeast.[160,162] The analogous mammalian protein has been localized by using epitope-tagged human *erd2*[163,164] and antibodies to the carboxy-terminal 21 amino acids of the bovine protein.[165] In normal cells, the ERD2 protein is found in the Golgi complex, but the protein is redistributed to the ER when the cells also express high levels of an appropriate ligand, such as KDEL-tagged lysozyme.[164] Thus, in mammalian cells as in yeast, the ERD2 receptor likely recycles together with the ER protein back to that compartment.[166] A cDNA clone similar in sequence and size to members of the *erd2* gene family has been identified in *Arabidopsis thaliana*.[167] The *Arabidopsis* protein exhibits 52% identity with the human homolog, and is 49% identical with the yeast counterpart. A functional role for the *Arabidopsis* ERD2 protein as a receptor for ER-retained (resident) proteins in plants is indicated by its ability to complement the lethal phenotype of the *erd2* deletion mutant of *S. cerevisiae*. In contrast, the human *erd2* homolog is unable to restore function in yeasts despite the 51% amino acid identity between the human and yeast proteins.[163] In fact, there appears

like the SEC62-SEC63 protein complex has also been suggested.[93]

Identification of potential components of the mammalian translocon has also been aided by the use of photochemicals or chemicals that cross-link nascent secretory proteins to adjacent integral ER membrane proteins. These studies implicated a 35- to 39-kDa glycoprotein (translocating chain-associated membrane [TRAM] protein) and a 34- to 37-kDa nonglycoprotein (homologous to yeast SEC61 and *Escherichia coli* SECY proteins).[114,115]

Reconstitution of detergent-solubilized rough ER proteins into translocation competent proteoliposomes allows the functional evaluation of putative translocation factors.[108,114] In mammalian cells, the SRP, the SRP receptor, SEC61, and the TRAM protein may participate directly in translocation because translocation is successful using these purified components. However, several additional stimulatory factors probably exist (perhaps mammalian homologs of the yeast proteins, SEC62, and SEC63). The TRAM protein may contact the N-terminal regions of nascent polypeptides at an early stage; a requirement for this protein depends upon the synthesized product. For example, it is essential for the transport of several precursors, but is only stimulatory for others. Further components assumed to be involved in the process include putative ribosome receptors of 34 kDa[116,117] and 180 kDa,[118] a membrane protein of 30 kDa with affinity for SRP (mp30),[119] and an unidentified ATP-binding membrane protein.[120] However, the function of these components, if any, is either disputed or not demonstrated yet (Figure 10) (see Reference 108 and references therein).

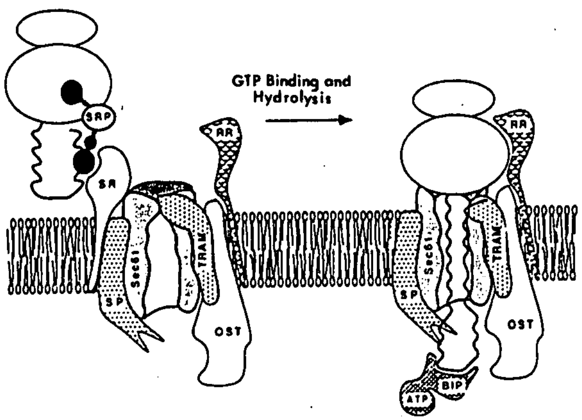


FIGURE 10. A schematic representation of the targeting and transport phases of a protein translocation reaction. The left portion of the diagram depicts SRP-dependent targeting of a ribosome to a preassembled translocation site comprised of the SRP receptor (SR), SEC61, TRAM, signal peptidase (SP), oligosaccharyltransferase (OST), and a ribosome receptor (RR). Conformational changes in Sec61p and TRAM that mediate channel gating are speculative. GTP binding and hydrolysis reactions permit SRP dissociation from the signal sequence and SRP receptor. Polypeptides traversing the channel contact TRAM and SEC61. Luminal reactions mediated by signal peptidase and oligosaccharyltransferase may inhibit reversible transport through the channel, whereas BiP may actively facilitate transport by binding the precursor as it emerges into the lumen. (From Gilmore, R., *Cell*, 75, 589, 1993. With permission from Cell Press.)

A hydrophilic protein is completely translocated across the rough ER if it is preceded by a signal that undergoes cleavage as translocation is occurring. However, in cases where the signal peptide is located internally, or its hydrophobic core is integrated within a longer membrane anchor domain, cleavage does not occur[15] (see later discussion). In this latter case, the "stop-transfer" sequence halts translocation and remains embedded in the membrane, functioning as the anchor for integral membrane proteins. It has been suggested that the proteinaceous ER translocation channel contains a receptor for stop-transfer sequences, the channel opening laterally to allow integration of membrane proteins.[93,102]

Protein synthesis (i.e., translation) and translocation of the protein generally occur simultaneously. Strong evidence argues against a mechanistic coupling between these two events; however, their temporal association may nonetheless play an important role — that of ensuring that nascent polypeptides maintain (or acquire) a translocation-competent conformation.[5] In mammalian cells, binding of the SRP to the signal peptide and the arrest of translation until functional insertion into the ER occurs would largely perform this important function,[15] although accessory proteins have also been implicated. SRP-like proteins have been detected in yeast. But this organism may also have a SRP-independent targeting pathway because it can survive without the SRP, and various proteins can be translocated *in vitro* in the absence of the SRP and the SRP receptor.[121] Any posttranslational translocation must occur without the SRP because the latter binds only to nascent chains attached to ribosomes.[122] It is likely that cytosolic chaperones such as 70-kDa heat-shock proteins (and probably additional proteins) function to assist proteins in maintaining a translocation-competent state.[107,123,124]

In most cases, the signal peptide is removed (in a cotranslational manner) from the nascent polypeptide chain as it emerges into the ER lumen by a signal peptidase located on the inner face of the membrane. Cleavage generally occurs at the more polar C-terminal region of the signal sequence and is thought to allow subsequent processes like protein folding and assembly to proceed normally (reviewed in Reference 4). For example, when the soybean storage protein glycinin is synthesized in an *in vitro* translation system (in the absence of ER membranes) from a synthetic mRNA corresponding to the entire glycinin coding region, the presence of the signal peptide inhibits assembly (trimerization) and results in misfolding of monomers.[125] The signal peptide has a similar negative effect on the acquisition of a functional (biologically active) conformation of α-amylase.[126]

b. Signal-Dependent Retention of Soluble Proteins in the ER Lumen

As mentioned previously, soluble resident proteins of the ER are distinguished by a carboxy-terminal tetrapeptide sequence. In mammals, these sequences are typically K/H/RDEL[59,70,127] and H/DDEL in yeasts.[127,128] Some of the more abundant soluble resident ER proteins of eukaryotes identified with these carboxy-terminal sequences include BiP, protein disulfide isomerase (PDI), and others.[129-131] In plants, the sequences have been identified at the carboxy-termini of several ER proteins (e.g., BiP, in tomato, maize, and tobacco; an auxin-binding protein in maize; and PDI in alfalfa).[17,59,132-142] The carboxy-terminal tetrapeptide sequence is necessary for retention of proteins in the ER.[70,143,144] When the sequence is deleted from an ER-resident protein (e.g., BiP), secretion is the outcome;[145] conversely, addition of the tetrapeptide to the carboxy-terminus of lysozyme (a protein that is normally secreted) or cathepsin D (a lysosomal enzyme) causes the proteins to accumulate in the ER.[70,143] Although some changes to individual amino acids can be tolerated, extending the sequence with random amino acids at the carboxy-terminus (or changing KDEL to KDAS) results in secretion of the target protein.[70,146] Thus, retention may be mediated by a specific receptor that recognizes (H/K/R)DEL only when it is present at the extreme carboxy-terminus of a protein (but see later discussion). In addition to KDEL, the carboxy-terminal sequences HDEL and RDEL are sufficient for retention in the ER and nuclear envelope of the reporter enzyme phosphinothricin acetyl transferase (PAT) in stably and transiently transformed plant cells

from a variety of species,[140,141] indicating that the plant ER retention mechanism can recognize several sequences. However, changing certain sequences within KDEL (such as KEEL, SDEL, KDDL, KDEI, and KDEV), leads to partial or complete loss of ER retention.

The nature of ER retention signals and the machinery involved in their recognition may be more complex than originally anticipated and may be distinct (in certain aspects) in different eukaryotic organisms. In addition to KDEL, HDEL, and RDEL, a variety of other sequences occur at the C-termini of ER-resident proteins in plants, yeast and mammals. Likewise, several other sequences are able to relocalize passenger proteins to the ER lumen in these eukaryotic organisms (Table 4) (reviewed in Reference 4; see references therein). Moreover, such unrelated sequences as the conserved *s*-cyclophilin-specific C-terminal region VEKPFAIAKE is sufficient to direct a secretory protein toward ER subcompartments.[147] In addition, there are examples of proteins that contain a carboxyterminal KDEL sequence (as deduced from their cDNA sequences), but are found in sites other than the ER lumen. An example is the plant auxin binding protein (ABP1) that is secreted (via the Golgi complex) to the cell surface (the plasma membrane and cell wall).[148] Interaction between the ER retention receptor and the KDEL tetrapeptide on this protein is suggested to be conditional. It may be regulated by auxin because binding of this hormone appears to change the structure of the protein (possibly its C-terminus and/or the accessibility of the KDEL retention signal), as revealed by an altered affinity of the auxin-ABP1 complex to monoclonal anti-ABP1-KDEL antibodies.[149] However, more information is needed to confirm these possibilities. PDI, containing a KDEL sequence at its C-terminal end, is not only found within the ER lumen but is also transported to the plasma membrane via the secretory pathway in rat exocrine pancreatic cells.[150] Thus, it is becoming increasingly apparent that the nature of ER localization signals is not obvious from the primary protein structure of ER-resident soluble proteins. Other approaches such as analyses of the crystal structures of R/H/KDEL-containing proteins[151,152] and nuclear magnetic resonance (NMR) analyses of a variety of peptides comprising the C-termini of ER resident proteins may be necessary to elucidate the critical features of ER localization signals.[4]

TABLE 4
Tetrapeptides Occurring in ER Proteins (A) and Comparative Analysis of Sequence Motifs Using Carrier Proteins (B)

	Mammalian cells	Plant cells
(A)	KDEL	KDEL
	RDEL	HDEL
	HDEL	NDELK
	KEEL	KQEL
	QEDL	
	HTEL	
	HIEL	
(B)	KDEL	KDEL
	RDEL	RDEL
	HDEL	HDEL
	KEEL	KEEL
	KEDL	SDEL
	KDDL	
	KNEL	
	REEL	
	REDL	
	RDDL	
	QDEL	
	QEEL	
	DKEL	
	KDEI	

From Vitale, A., Ceriotti, A., and Denecke, J., *J. Exp. Bot.*, 44, 1417, 1993. With permission from Oxford University Press. See references therein.

The addition of KDEL or HDEL onto the carboxy-terminus of some proteins retards transport, but is not sufficient for their absolute retention in the ER.[153,154] Retardation of export rather than total ER retention occurs with several mammalian secretory proteins that are modified at the exact carboxy-terminus to encode the sequence SEKDEL.[153] When the carboxy-terminal sequence of a plant vacuolar protein (PHA) is changed from LNQIL to LNKDEL and the derivative protein expressed in transgenic tobacco, a large proportion is localized in the ER and the nuclear envelope (that is continuous with the ER). However, a considerable amount of the PHA-KDEL is transported through the Golgi complex and reaches the protein storage vacuole/protein body.[154] Partial retention of these proteins in the ER may be due to a less than optimal display of the carboxy-terminal KDEL sequence, thus diminishing its

recognition by the appropriate receptor[155] and allowing its escape from the normal retrieval mechanism (see subsequent discussion). Other structural features of the protein may be required for ER retention, and their role may not be restricted to simply ensuring optimal positioning of the tetrapeptide for receptor-mediated recognition.[4,17]

Changing the carboxy-terminus of the vacuolar pea storage protein vicilin to include KDEL (SEKDEL) results in a dramatic increase in the accumulation of vicilin in the leaves of transgenic tobacco and alfalfa.[156,157] The retention signal keeps the protein from advancing to a compartment where it could be degraded (presumably the vacuole) and results in the formation of protein-body-like inclusions. These ER-derived, membrane-bound, electron-dense structures (0.5 to 1.0 µm in diameter) contain vicilin and resemble the ER-derived protein bodies found in the endosperm cells of certain cereals such as maize and sorghum (see Section V.G.1).

Given that most proteins are transported by bulk flow, whereas certain proteins are selectively retained, it is pertinent to examine the mechanism of specific retention in the ER. The luminal proteins are not associated with a membrane-bound receptor,[59,145] nor is retention guaranteed by membrane anchoring; many membrane-bound viral proteins are transported out of the ER.[60] It appears that for soluble proteins at least, ER retention mechanisms come into play at a later point in the secretory pathway, that is, after the proteins have been transported out of the ER.[59,143] More specifically, luminal ER proteins that escape are continuously and specifically retrieved from a post-ER ("salvage") compartment (that is pre-Golgi or within the Golgi complex) by a receptor;[158] the receptor-ligand complexes are cycled back to the ER (by specialized vesicular transport; see subsequent discussion), where the KDEL-bearing proteins are released.[59] In yeast (*Saccharomyces cerevisiae*), the retrieval mechanism appears to be in the Golgi complex,[127,159] and some very elegant studies in this organism[160–162] have identified a putative receptor involved in HDEL recognition. A genetic approach was utilized to identify the essential components of the "recycling" pathway; to this end, *S. cerevisiae* mutants specifically defective in the HDEL-mediated ER retention system were isolated, allowing subsequent identification of two genes required in the process, viz., *erd1* and *erd2* (for "ER retention defective"). Deletion of the *erd1* gene causes a pleiotropic defect in part of the Golgi apparatus, which evidently results in inefficient retrieval from this organelle.[161] The *erd2* gene encodes a 26-kDa integral membrane protein that is required both for retention of ER proteins and perhaps indirectly for normal traffic of proteins through the Golgi.[160] Strikingly, the abundance of this protein determines both the efficiency and capacity of the retention system: reduced expression leads to secretion of HDEL-tagged proteins; conversely, its overexpression improves their retention (both in wild-type cells and in other mutants). Moreover, *erd2* determines the signal specificity of the retention system;[162] exchange of the *erd2* gene from *S. cerevisiae* (which only recognizes HDEL) for the corresponding gene from another yeast, *Kluyveromyces lactis* (which recognizes either HDEL or DDEL), allows equal recognition of DDEL and HDEL in *S. cerevisiae*. These observations, as well as the subcellular localization of the protein (i.e., in a post-ER, Golgi-like compartment), make it an excellent candidate for the receptor that sorts lumenal ER proteins in yeast.[160,162] The analogous mammalian protein has been localized by using epitope-tagged human *erd2*[163,164] and antibodies to the carboxy-terminal 21 amino acids of the bovine protein.[165] In normal cells, the ERD2 protein is found in the Golgi complex, but the protein is redistributed to the ER when the cells also express high levels of an appropriate ligand, such as KDEL-tagged lysozyme.[164] Thus, in mammalian cells as in yeast, the ERD2 receptor likely recycles together with the ER protein back to that compartment.[166] A cDNA clone similar in sequence and size to members of the *erd2* gene family has been identified in *Arabidopsis thaliana*.[167] The *Arabidopsis* protein exhibits 52% identity with the human homolog, and is 49% identical with the yeast counterpart. A functional role for the *Arabidopsis* ERD2 protein as a receptor for ER-retained (resident) proteins in plants is indicated by its ability to complement the lethal phenotype of the *erd2* deletion mutant of *S. cerevisiae*. In contrast, the human *erd2* homolog is unable to restore function in yeasts despite the 51% amino acid identity between the human and yeast proteins.[163] In fact, there appears

to be several distinctive features of the ER-retention mechanisms of yeast (e.g., *S. cerevisiae*) vs. those of mammals.

In *S. cerevisiae*, there is a strict requirement for a C-terminal HDEL sequence for ER retention. In animals and plants, many different C-terminal sequences on ER-resident proteins are tolerated, such as HDEL, RDEL, and KDEL (Table 4; see earlier discussion). Moreover, the monoclonal antibody ID3 (which recognizes the C-terminal portion of mammalian PDI) does not recognize HDEL-containing yeast BiP,[155] nor any other yeast proteins.[4,168] However this antibody is capable of recognizing an important structure of the ER localization signal in a family of plant and mammalian ER-resident proteins. In some ways then, the ER retention mechanisms of plants exhibit characteristics somewhat intermediate between those of animals and those of yeast.

The plant ERD2 protein may be an example of a functional link between the other two groups of eukaryotic organisms. A binding specificity similar to the human ERD2 counterpart is suggested and it contains conserved sequences (e.g., Asp[176]) found to be important for recycling of the human receptor (Asp[193]). However, it can also function in yeast.[167] It is expected that the differences between the *erd2* homologs from plants and animals with respect to their functional activity in yeast are a result of differences in critical amino acid sequences of the two proteins. Future prospects for mapping the functional domains of the ERD2 proteins include mutational analyses and construction of chimeric proteins using the different *erd2* genes in conjunction with complementation tests, binding, and localization studies.[167]

Another pertinent question that has recently been addressed is how the ER membrane is maintained in the face of a heavy flow of vesicular traffic towards the Golgi stack.[169,170] The ER is a dynamic network of tubular membranes that extend along microtubules, continuously breaking, fusing, and rearranging.[171] Recent studies utilizing the antifungal antibiotic brefeldin A suggest that there is an intermediate compartment to the ER transport pathway involving microtubules that potentially could restore membrane components and escaped proteins to the ER.[169,170] In the presence of the drug, the Golgi complex breaks down and appears to merge with the ER;[172-174] consequently, both components now interact with microtubules in the same way. Secretory and (H/K)DEL-tagged proteins may normally be sorted in the intermediate (salvage) compartment, such that the latter are returned via a retrograde pathway involving microtubules. Membrane vesicles destined to return to the ER may carry or acquire proteins that enable them to interact with microtubules; these proteins may normally be excluded from vesicles involved in nonselective (bulk-flow) transport during sorting in the salvage compartment.[15,169,170] Because the retrograde pathway into the ER depends on microtubules, in contrast to the anterograde pathway out of the ER, this provides a potential means by which cells can regulate membrane transport along these two pathways (Figure 11).

The ERD2 protein in yeast not only functions as a receptor, but also is required for normal vesicular traffic through the Golgi complex; these two functions may be related.[160] For example, ERD2 may be required for the formation of the specialized vesicles involved in retrograde transport; whether it interacts directly with microtubules remains to be determined. In mammalian cells, proteins localized in a pre-Golgi (or Golgi) intermediate compartment may play a role similar to that of ERD2 (e.g., as microtubule attachment proteins). Some (but not all) ER membrane proteins normally escape to the Golgi complex and require retrieval; recycling of the SEC12 protein (see later discussion) and perhaps other membrane proteins involved in vesicular traffic and membrane flow is also suggested.[175] ERD2 and equivalent mammalian proteins might also be involved in maintaining the supply of such components in the compartments from which vesicles bud. A failure to retrieve these membrane proteins could reduce the flow of the secretory pathway to a trickle.[160] Whether a similar retrograde pathway and retrieval mechanism exist in plant cells remains to be determined. Studies using brefeldin in plants are in their infancy and we clearly need to know more about the specific effects of this inhibitor on the plant Golgi complex and ER both at the morphological level and at the molecular level.[176,177]

Brefeldin inhibits the secretion of proteins[148,178] and noncellulosic polysaccharides[179] and the transport of a vacuolar protein, PHA.[180] Evidence supporting a retrograde pathway from the Golgi complex to the ER is indicated by changes in the quantity of various glycosylated forms of PHA in transgenic tobacco cells following treatment with brefeldin that appear to be consistent with a redistribution of Golgi enzymes involved in glycan processing to the ER.[180] However, redistribution of the Golgi membrane marker JIM84 into the ER does not occur in maize and onion root cells,[3,177,181] and immunofluorescence studies fail to show tubular structures emerging from the Golgi complex and directed to the ER. Several features of the plant Golgi complex make it unique and distinct from the Golgi of animals, including separate stacks, wall matrix synthesis, and lack of fragmentation during plant cell division. There even may be structural and functional cell type-specific differences in the organization of the plant Golgi stacks,[83] and these are reflected in a differential sensitivity to brefeldin (see Reference 177 and references therein). It is apparent that more systematic studies need to be undertaken before conclusions can be reached regarding a retrograde transport pathway from the plant Golgi complex to the ER.

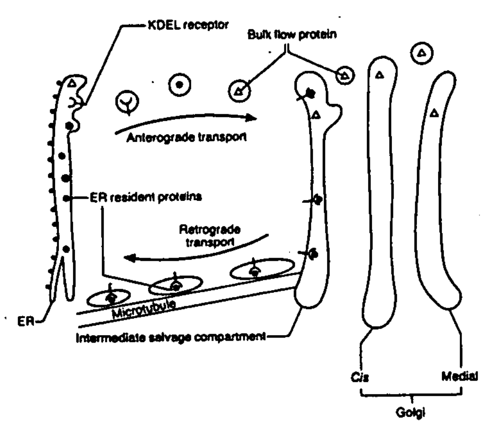


FIGURE 11. Protein recycling between the ER and Golgi. Proteins are transported in vesicles from the ER to the salvage compartment, where the KDEL receptor carries proteins bound back to the ER by a retrograde transport pathway in which vesicles travel along microtubules. (From Austen, B. M. and Westwood, O. M. R., *Protein Targeting and Secretion*, Oxford University Press, Oxford, 1991. With permission from Oxford University Press.)

c. Targeting Signals for Membrane Proteins of the Secretory Pathway

I. Topology of Integral Membrane Proteins

The membranes surrounding the organelles of the secretory pathway represent highly specialized compartments; most of the functions of membranes are carried out by the select group of proteins embedded within, or otherwise associated with, the lipid bilayer. (For further discussion on the topography of membrane proteins, see References 10 and 97). After translocating partly across the ER membrane, these proteins are transported to further compartments of the secretory pathway (along with luminal proteins) in the membranes of the transport vesicles. Transmembrane proteins are either single-pass (i.e., they span the

phospholipid bilayer only once) or multi-pass (i.e., they span the bilayer multiple times). Topology is established when the protein first assembles into the membrane. For plasma membrane proteins and membrane proteins of internal organelles of the secretory pathway, topology is adopted at the ER membrane; subsequent transport to other organelles or to the cell surface is mediated by vesicles in which protein topology remains unchanged. Thus, a domain of a membrane protein that remains cytoplasmic at the ER membrane is still cytoplasmic after transport of the protein to the organelle or plasma membrane. Proteins are also specifically oriented within the membranes of mitochondria, chloroplasts, and other organelles after synthesis in the cytoplasm.[10]

The orientation of single-pass transmembrane proteins depends on the positions of the signal peptide and the hydrophobic stretches of the protein that will span the membrane after the protein is oriented. As shown in Figure 12 and Table 5, membrane proteins are classified into three groups (I, II, and III) based primarily on how they are oriented in the membrane and whether they contain a (conventional) signal peptide. Not all membrane proteins pass through the lipid bilayer; some become attached to membranes by a specific covalent attachment of fatty acids, prenyl groups, or inositol glycolipids.[10,94]

ii. Localization Signals of Membrane Proteins

Proteins destined to reside in the membranes of the ER cisternae require retention signals; however, much less is known about the nature of these ER membrane localization signals. The C-terminal KDEL/HDEL signal used for the retrieval of lumenal ER proteins from the Golgi complex (salvage compartment) is present on the lumenal domains of two yeast type II membrane proteins (the products of the *sec20* and *sed4* genes) and appears to be involved in their retrieval back to the ER.[182] However, other retention mechanisms exist. Some mammalian ER membrane proteins (e.g., glucuronosyl transferase) contain a double lysine motif in the C-terminal cytoplasmic tail (KKXX, KXKXX, or a similar sequence), which functions to ensure retrieval from the Golgi complex back to the ER.[183] A double lysine motif is also present in the C-terminal sequence of a plant gene, omega-3 desaturase from *Arabidopsis*.[184] The E19 protein of adenovirus is a membrane-anchored resident ER polypeptide that has a short tail protruding into the cytoplasm. A small stretch of amino acids on this exposed tail (i.e., the last C-terminal six residues, DEKKMP) is necessary and sufficient for retention in the ER;[72] the last three amino acids are conserved in E19 proteins of several adenoviruses.[185] However, this sequence is not conserved in other ER membrane proteins. Other positively charged amino acids within the C-terminus might constitute retention signals for ER membrane proteins, as indicated by the C-terminal sequence of calnexin (SPRNRKPRRE), an abundant ER type I membrane protein of mammalian cells.[186] However, the ER retention signal of the CD3ε chain of the T-cell receptor appears to be completely unrelated to the lysine motif.[187] Still other retention signals have been identified, primarily of viral proteins in infected animal cells. For the rotovirus VP7 protein (a 37-kDa glycoprotein), the signal peptide, together with the first 60 amino acid residues of the mature protein, forms the retention signal,[188] although the mechanism of retention is not yet understood. A N-terminal amino acid sequence is also responsible for the retention of a hepatitis B virus type II membrane glycoprotein.[189] Thus, the mechanisms for localization of ER membrane proteins are diverse, and no obvious consensus sequence can be defined. Moreover, both cytoplasmic and luminal domain recognition is involved. However, it is clear that a retention mechanism operating solely by salvaging lost or escaped proteins is not sufficient to explain the available evidence. For example, careful examination of the oligosaccharides of some ER membrane proteins indicates that they do not leave the ER.[190,191] Furthermore, removal of the retrieval signal from ER proteins often causes them to be transported only very slowly, suggesting that there is additional retention information elsewhere in the protein.[70,183] Surprisingly, a KKXX sequence is also found in ERGIC 53, a marker protein for an intermediate compartment between the ER and Golgi complex.[192] Presumably, this protein lacks the necessary structural determinants for ER retention, causing its rapid export from the ER.

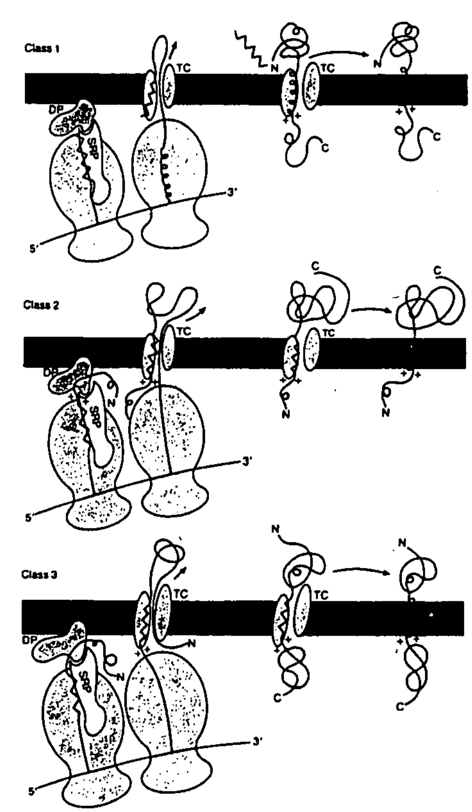


FIGURE 12. Mechanisms of assembly of single transmembrane proteins. The three different modes of assembly for proteins involve interactions of the signal sequence or stop-transfer sequence with signal recognition particle (SRP), docking protein (DP), and the translocation complex (TC). (From Austen, B. M. and Westwood, O. M. R., *Protein Targeting and Secretion*, Oxford University Press, Oxford, 1991. With permission from Oxford University Press.)

Retention signals for localization in the Golgi membrane have been identified for several *medial-* and *trans-*Golgi enzymes. They are all type II proteins with a short N-terminal cytoplasmic tail followed by a single transmembrane domain (reviewed in Reference 7). This transmembrane domain (with a few amino acids on either side) is both necessary and sufficient for Golgi localization. The mechanism is highly specific; when the membrane spanning domain of a reporter protein is replaced by that of the Golgi protein, the hybrid protein not only resides in the Golgi complex, but

TABLE 5
Topography of Single-Pass Transmembrane Proteins and Their Mechanisms of Insertion Into the Membrane[a]

Type	Examples	Characteristics
I	Glycophorin A VSV-G M6P receptor	Signal sequence interacts with SRP; subsequent stop-transfer sequence halts translocation; signal sequence cleaved; oriented with N-termini on luminal side; C-termini on cytoplasmic side
II	Ca^{2+}-ATPase Lens MP26	Lacks cleavable signal peptide; stop-transfer sequence doubles as signal sequence that interacts with SRP and enters translocation site; have higher concentration of positively charged residues on N-terminal side of stop-transfer sequence C-termini reside in lumen, N-termini on cytoplasmic side
III	Cytochrome P-450 Epoxide hydrase	Mechanism of insertion as above; have a number of positively charged residues on C-terminal side of stop-transfer sequence; N-termini in lumen; C-termini in cytoplasm

[a] Reviewed in Reference 10; see also Figure 13. (VSV-G: G-protein of vesicular stomatitis virus; M6P receptor: mannose-6-phosphate receptor.)

also in the correct cisternae.[193-195] In contrast to the proteins of the TGN (see later discussion) there is little evidence that the enzymes of the Golgi stack recycle through later compartments. On the other hand, a resident of the TGN, TGN38, contains two non-overlapping signals, one mediating retrieval and the other mediating retention.[196] A tyrosine-based motif in the cytoplasmic domain acts as the retrieval signal, returning the protein from the cell surface to the TGN. The second signal is within the membrane spanning domain and is necessary for retention in the TGN.

Selective retention mechanisms mediated by transmembrane domains are likely to involve some interaction within the membrane itself. One possibility is that retention results from the formation of protein aggregates (reviewed in Reference 7). For example, some of the *medial*-Golgi enzymes (*N*-acetylglucosaminyl transferase I and mannosidase II) can associate with each other *in vivo*, suggesting that they may form complexes that are too large to be transported. Retention could be improved further by binding of the proteins to an intercisternal matrix that would effectively exclude them from regions where vesicle buds form.[194] However, whereas the transmembrane domain is sufficient for retention, it does not appear to be sufficient for aggregation (i.e., the structural requirements for retention and complex formation are somewhat different). Aggregation or matrix attachment, then, may not be the primary retention mechanism for Golgi proteins. However, such interactions may well improve retention efficiency and ensure that enzymes required in the same compartment are co-localized; they may also play a role in stabilizing and organizing the stacked structure of the Golgi complex.[7,194] An alternative view is that the retention of Golgi enzymes is a consequence of the general physical properties of the transmembrane protein domains and the membrane in which they lie.[197] Here, the length of the transmembrane domain, rather than any specific amino acid sequence, is the critical feature for sorting. The membrane-spanning domains of Golgi proteins are consistently shorter (typically 17 residues long), and have a greater proportion of phenylalanine residues than those of plasma membrane proteins (typically 21 residues long). Moreover, replacement of the transmembrane domain of sialyltransferase, a Golgi-specific protein, with 17 leucine residues does not interfere with its retention in the Golgi complex, but extension of the replacement amino acid sequence to 23 leucines allows transport to the cell surface.[198] Thus, a model has emerged that views the sorting of Golgi

proteins as a consequence of the sorting of membrane lipids, with Golgi and plasma membrane proteins adapted to bilayers of different thicknesses. Furthermore, the polarity of the Golgi complex is thought to arise naturally from a gradient of lipid composition; the stable identity of this organelle being maintained by a simple feedback loop: protein content controls lipid composition, which in turn controls protein content. Before this model can be substantiated, more information is needed concerning the mechanisms and consequences of lipid sorting as well as targeting molecules involved in vesicular targeting[7] and selection of vesicle content (see later discussion).

There is information on the topogenic sequences of internal membrane proteins that are incorporated into the membranes (ER, Golgi, tonoplast, plasma membrane) of the secretory system of plant cells. Analysis of the amino acid sequences of three proteins that transit the secretory pathway (e.g., a plasma membrane proton-translocating ATPase, a vacuolar H^+-pyrophosphatase, and a seed-specific tonoplast protein) reveals several membrane-spanning domains in each protein.[199-202] The tonoplast intrinsic protein (TIP) of bean seeds (α-TIP) is synthesized on the rough ER, and its transport to the tonoplast is mediated by the secretory system; its C-terminal transmembrane domain may contain sufficient information for transport to the tonoplast.[203] However, the possibility that the signal-independent default pathway for membrane proteins leads to the tonoplast (rather than to the plasma membrane) is being investigated (see later discussion).[180]

Interestingly, synthesis of TIP in seeds does not appear to be related (in a quantitative manner) to storage protein deposition, and TIP is not found in leaves that accumulate vegetative storage protein. A role as a solute transporter in seeds has been suggested for this protein;[200] alternatively, its developmental regulation is indicative of a physiological function during late seed maturation, perhaps in the acquisition of desiccation tolerance.[204] Recently, the function of γ-TIP has been elucidated — that of forming water channels involved in osmoregulation.[39] The vacuolar H^+-ATPases are thought to be an integral part of the organelles of the endomembrane system in plants (reviewed in References 205 through 207); acidification of these compartments by the enzyme may be essential for protein sorting in plants, similar to its role in animal and yeast cells (see later discussion).

C. Membrane Growth and Lipid Transfer

As mentioned previously, different types of plant cells contain different proportions of various membranes (reviewed in Reference 50), and this has raised questions about the mechanisms that control membrane growth and changes in membrane lipid content/composition, requiring the movement of lipids between organelles. There may be two mechanisms involved in the selective transfer of lipid between organelles of the secretory pathway (e.g., the ER, Golgi, and plasma membrane), resulting in different membrane lipid compositions.[50] These include the transfer of lipid in the membranes of ER-derived vesicles that fuse with the acceptor membrane[66] and the transfer of lipid monomers through the cytoplasm as either free monomers or via protein-facilitated transport. Lipid-transfer proteins (LTPs) were originally defined by their ability to transfer specific or several different phospholipids between membrane fractions *in vitro*, but their *in vivo* function(s) is poorly understood.[208] A LTP in yeast (the phosphatidyl-inositol-specific LTP, or pi LTP) may be involved in transferring lipid from the Golgi complex back to the ER; a temperature-sensitive mutant deficient in this protein (and in the formation of secretory vesicles) accumulates very large amounts of Golgi membrane, and the ER virtually disappears.[209] A role for the nonspecific LTPs of plants as catalysts for a similar transfer of lipid between membranes is far from established. For example, the subcellular location of one such protein, a nonspecific LTP of *Arabidopsis* that is localized to the cell wall,[210] is inconsistent with a role in intracellular lipid transfer. Analysis of transgenic plants in which the amount of nonspecific LTP has been greatly reduced by expression of an antisense gene construct[211] may aid in elucidating the *in vivo* function of plant nonspecific LTPs.[50]

If nonspecific LTPs play a role in intracellular lipid transfer, what is their mechanism of action? One possibility is that these proteins

preferentially abstract lipid from a membrane of high lipid/protein ratio and release lipid into membranes with a low lipid/protein ratio.[50] It is further speculated that membrane growth is controlled via a mechanism that enables cells to sense the lipid/protein ratio of its membranes. Integral membrane proteins, which are often specific to a particular membrane and are regulated developmentally and by environmental cues, have been pinpointed as the important sensor mechanism. According to this model, then, membrane growth is triggered by the synthesis of membrane proteins. Insertion of protein into existing membranes triggers the demand for lipid that is either synthesized within the membrane or transferred from the ER by a nonspecific LTP or an equivalent mechanism. Other regulatory factors to consider are protein turnover and the possibility that changes in the amount of protein may induce changes in the composition of various lipids.[50]

D. Cotranslational and Posttranslational Modifications Along the Secretory Pathway

Many proteins that are transported along the secretory pathway in plant cells undergo extensive modifications, including glycosylation, folding, oligomer assembly, and proteolytic processing. Some of these processes have been studied in relation to their role in transport efficiency, protein stability, and protein targeting per se.

1. Glycosylation

Glycosylation is a modification that many vacuolar and extracellular proteins undergo *en route* to their target organelles. In the process known as O-linked glycosylation, the oligosaccharide or glycan of the mature glycoprotein is linked to the oxygen of either a serine or threonine amino acid residue. In N-linked glycosylation, the glycan is linked to the amide nitrogen of an asparagine residue. O-linked glycans are important components of hydroxyproline-rich glycoproteins and arabinogalactan proteins, two major classes of cell-wall proteins (see earlier discussion). In addition to O-glycosylating serine and theonine residues, plants frequently attach oligomeric arabinoside chains to hydroxyproline residues of proteins; both these modifications are carried out in the Golgi complex.[212,213] As in animal systems, the O-linked glycans of cell-wall proteins may be of critical importance for producing stiff and extended protein conformations.[214,215]

In their mature state, vacuolar or extracellular and membrane glycoproteins often have both high mannose and complex N-linked glycans (Figure 13) (reviewed in References 17, 29, and 216 through 219). Glycans of the high-mannose type (e.g., $Glc_3Man_9GlcNAc_2$) are assembled on dolichol pyrophosphate lipid carriers in the ER lumen. These are transferred (in a cotranslational manner) to specific Asn residues (of the acceptor sequence Asn-X-Ser/Thr) on nascent polypeptides by oligosaccharyl transferase[220] (reviewed in Reference 221). Two ER-resident enzymes (Glucosidases I and II) subsequently remove the three terminal glucose residues from the high-mannose glycans on most proteins[222] (a notable exception is α-mannosidase of jackbean cotyledons, which escapes this modification).[223]

The dolichol pyrophosphate lipid carrier is embedded in the lipid bilayer, and the high-mannose glycan is built up by the stepwise addition of individual monosaccharides using sugar nucleotides as the donors. In mammalian cells and yeast, stepwise addition of sugars to yield the structure $Man_{3-5}GlcNAc_2$-PP-dolichol is thought to occur on the cytosolic face of the ER membrane; following translocation of the entire intermediate to the luminal face, the subsequent addition of sugars is completed (Figure 14) (reviewed in Reference 224). The product, dolichol PP-$GlcNAc_2Man_9$ is formed by the addition of four mannose residues on the luminal side; here, the direct sugar donors are not guanosine diphosphate (GDP)-Man, (a membrane-impermeable substrate), but dolichol-P-mannose, the latter being synthesized on the cytosolic side of the membrane from GDP-Man and then translocated into the lumen. Dolichol-PP-$GlcNAc_2Man_9$ must be glucosylated before it can serve as an oligosaccharide substrate for nascent polypeptide chains; the donor for this reaction is dolichol-P-glucose, which is itself synthesized from UDP-Glc and dolichol-P. The catalytic subunit of the plant enzyme Glc-P-dolichol synthase (a polypeptide of

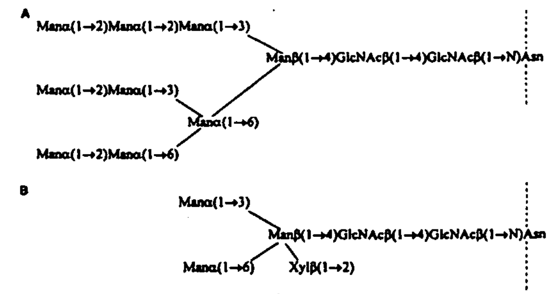


FIGURE 13. The glycan side chains of glycoproteins. (A) A typical Asn-linked simple or high mannose oligosaccharide side chain (soybean agglutinin). (B) Probable structure of one version of a modified or complex oligosaccharide side chain of phaseolin. (From Bewley, J. D. and Black, M., *Seeds. Physiology of Development and Germination*, 2nd ed., Plenum Press, New York, 1994. With permission from Plenum Press.)

39 kDa) has been purified from mung bean microsomes.[225]

The enzyme oligosaccharyl transferase that is responsible for transfer of the high-mannose chain to the Asn residue of the growing polypeptide chain is thought to be part of an integral membrane protein complex, possibly comprised of ribophorins I and II and a polypeptide of 48 kDa.[226] One complex occurs at each protein translocation site (translocon) in the ER membrane; this may play a role in maximizing the efficiency (and accuracy) of glycosylation, allowing rapid continuous monitoring of the nascent and growing polypeptide chain for the presence of glycosylation sites.[4,226] As mentioned previously, elaboration of the glycan moiety to the protein normally occurs cotranslationally. The coupling of glycosylation with protein synthesis/translocation ensures an efficient transfer of the oligosaccharide as a result of the close proximity of the transferase complex in the ER membrane. Also, as will be discussed, glycosylation influences protein conformation. Thus, it may be important that glycosylation occur early in the pathway of translocation to ensure that the protein will subsequently fold and assemble properly, rendering it competent for transport out of the ER lumen (see later discussion).

Presence of the consensus tripeptide sequence Asn-X-Ser/Thr within a polypeptide is not sufficient to guarantee glycosylation; in many cases, the site is not glycosylated or is glycosylated with low efficiency (reviewed in Reference 4). As an example, the bean storage protein phaseolin has two glycosylation sites; the one closest to the N-terminus is always glycosylated, but utilization of the second site is not efficient.[218,227] The pattern of glycosylation must be characteristic of the protein because it generally does not depend on the cell type (or host organism) in which it is synthesized. When phaseolin is expressed in a variety of heterologous hosts (e.g., yeast, *Xenopus* oocytes, or tobacco protoplasts) it is produced as a mixture of partially and fully glycosylated polypeptides.[228-230] It is likely that the cotranslational folding of the polypeptide luminal parts (i.e., those already translocated) determines whether a subsequent site in the polypeptide chain will be a substrate for glycosylation by restricting access of the oligosaccharyl transferase to certain glycosylation sites. Supporting evidence comes from studies using agents that negatively affect folding; under these conditions, a site normally unused or inefficiently utilized *in vivo* becomes fully glycosylated.[231] Likewise, destruction of the

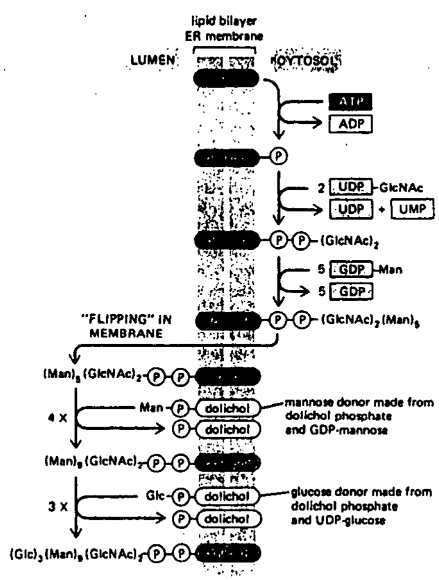


FIGURE 14. Synthesis of the lipid-linked oligosaccharide that is transferred to asparagine residues of nascent polypeptides on the luminal side of the ER membrane. The oligosaccharide is assembled sugar by sugar onto the carrier lipid dolichol (a polyisoprenoid), to which the first sugar is linked by a pyrophosphate bridge. This high-energy bond activates the oligosaccharide for its transfer from the lipid to an asparagine side chain. The synthesis of the oligosaccharide starts on the cytosolic side of the ER membrane and continues on the luminal face after the Man_5-$GlcNAc_2$ lipid intermediate is flipped across the bilayer. All of the glycosyl transfer reactions on the luminal side of the ER involve transfers from dolichol-P-glucose and dolichol-P-mannose. These activated, lipid-linked monosaccharides are synthesized from dolichol phosphate and UDP-glucose or GDP-mannose (as appropriate) on the cytosolic side of the ER, and are then thought to be flipped across the ER membrane. Abbreviations: GlcNAc, *N*-acetylglucosamine; Man, mannose; Glc, glucose. (From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D., *Molecular Biology of the Cell*, 2nd ed., Garland Publishing, New York, 1989. With permission from Garland Publishing.)

first glycosylation site of phaseolin by site-directed mutagenesis increases the efficiency of glycosylation at the second site *in vivo*.[230] Thus, an oligosaccharide chain on the portion of the polypeptide already synthesized can affect cotranslational glycosylation, either by directly masking a second subsequent glycosylation site or, indirectly, through an effect on folding.[4]

Interestingly, when a phaseolin mutant that is defective in its assembly into trimers is synthesized in *Xenopus* oocytes, it is retained in a pre-Golgi compartment (unlike the wild-type protein) and undergoes posttranslational glycosylation at the second site.[229] Posttranslational glycosylation of the ER-retained mutant may have occurred due to its prolonged residence in the ER and a better exposure of the consensus sequence in monomers than in trimers. There is likely a complex interplay between folding, assembly, and glycosylation to ensure that a newly synthesized protein acquires a functional conformation after translocation into the ER lumen.[4]

Following cotranslational transfer of the glycan onto the nascent protein, removal of the three Glc residues in the ER results in the formation of a of typical high-mannose glycan ($Man_9[GlcNAc]_2$). This may be further modified later in the secretory pathway (in the Golgi complex) (i.e., converted from a high-mannose glycan to a complex glycan as a result of the sequential action of several glycosidases and glycosyl transferases).[216,217] The complex glycans of plant glycoproteins differ substantially from those of mammalian glycoproteins; most notable is the lack of sialic acid and the presence of a xylose residue β-1,2 linked to β-mannose of the core. However, plant complex glycans may share some similar structural features with those of insects and molluscs.[232] In plants, complex glycans are characterized by fucose and xylose sugar residues attached to the proximal GlcNAc and core mannose residues, respectively. This modification renders the glycoprotein resistant to the *Streptomyces* enzyme endoglycosidase H; hence, this characteristic has been used as evidence for passage through the Golgi complex.[17]

As in animal cells, the posttranslational processing of N-linked glycans of plant proteins occurs in a *cis*-to-*trans* direction.[213,233] Progress has been made toward characterizing and localizing enzymes involved in converting N-linked glycans to complex glycans.[213] However, a clear-cut biochemical fractionation of *cis*-, *medial*-, and *trans*-Golgi cisternal membranes from plant cells has not been obtained to date.[222,234] To overcome some of the limitations associated with biochemical fractionation, immunolabeling techniques, in conjunction with quantiative electron microscopy, have been used to determine in which Golgi cisternae specific sugars are added to glycoproteins and polysaccharides.[235,236] This approach became feasible with the development of methods for the immunolabeling of high-pressure frozen/freeze-substituted cells in which the morphology of the Golgi stacks is preserved to a much greater extent than in conventionally fixed samples. Antibodies specific for the xylose residues β1,2-linked to mannose and for fucose α1,3-linked to GlcNAc have localized enzymes involved in glycan processing in suspension cultured sycamore cells. β1,2-xylosyl transferase is located mainly in *medial*-Golgi cisternae, whereas the α1,3-fucosyl transferase is confined primarily to the trans-Golgi cisternae and the TGN, indicating that addition of fucose occurs after addition of xylose.[213] Further biochemical and immunocytochemical studies (with antibodies specific to the sugars of complex glycans) will help to elucidate the compartmentalization and sequence of the processing pathways in the Golgi complex.

Apparently, glycans must be readily accessible to the appropriate enzymes in the Golgi apparatus; when inaccessible, they remain in a high-mannose form as the protein moves through the Golgi complex.[237] A large variety of complex structures may be formed on mature glycoproteins and high-mannose chains and different complex chains may be found on the same polypeptide.[238-240] Moreover, the final structure of a complex chain may be cell-type dependent due to some cell-specific differences in the set of Golgi processing enzymes.[4,239] In general, the complex glycans of vacuolar proteins tend to be very similar (Figure 13), and are closely related to ($[Xyl]Man_3[Fuc]GlcNAc_2$), whereas those of secreted proteins are very heterogeneous and fairly large;[239] these differences are apparent even in the same plant cells. The larger oligosaccharides have additional sugars, such as *N*-acetylglucosamine, galactose, and fucose, attached to the nonreducing ends of the molecules.[241,242] Some of the heterogeneity is generated by the action of glycosidases located at sites beyond the Golgi complex (e.g., in the cell wall). In carrot suspension cultured cells, a 52- to 54-kDa medium protein is secreted with large homogeneous complex glycans; the heterogeneity originates from slow processing after secretion. In the same cells, the complex glycans of

cell wall β-fructosidase are processed before the enzyme is integrated into the cell wall.[239] Vacuolar proteins leave the Golgi complex with one or two terminal *N*-acetylglucosamine residues that are removed soon after arrival at their final destination, giving rise to homogeneous complex glycan structures.[218,243]

What are the general roles of N-linked glycosylation and glycan processing? The role of N-linked glycans in the transport and targeting of proteins to the plant cell vacuole has been examined (reviewed in Reference 17), and the available evidence indicates that glycans do not contain vacuolar targeting information. Not all vacuolar proteins are glycoproteins (e.g., 11S globulins, legumin, glycinin, and others); moreover, some glycosylated precursors lose their glycans during posttranslational processing *en route* to the vacuole (e.g., concanavalin A, wheat germ agglutinin, and β-glucanase).[244-246] Tunicamycin inhibits N-linked glycosylation in the ER by interfering with the elaboration of glycans (GlcNAc-1-P) onto the dolichol lipid intermediate.[247] However, this antibiotic generally does not cause missorting of (normally glycosylated) vacuolar proteins, including pea vicilin and bean PHA.[248,249]

Elimination of glycosylation sites by site-directed mutagenesis has been used as a direct approach to study the role of glycosylation in protein transport. Elimination of the glycosylation sites of three plant vacuolar proteins by this means (viz., PHA, patatin, and barley lectin) does not lead to mistargeting of the unglycosylated proteins in a heterologous host plant system.[82,250,251] Thus, in contrast to proteins of the animal lysosome, in which glycans play a pivitol role in their targeting (see earlier discussion), the glycans of plant vacuolar proteins do not contain targeting information. Such is also the case for glycans of yeast vacuolar proteins.[252]

Whereas tunicamycin inhibits the biosynthesis of the N-linked glycans resulting in nonglycosylated proteins, a second class of inhibitors (e.g., the alkaloids castanospermine, deoxymannojirimycin, and swainsonine) inhibit the processing of N-linked glycans.[253,254] More specifically, castanospermine inhibits glucose trimming in the ER; in relation to trimming reactions involved in the conversion of glycans to the complex form in the Golgi, deoxymannojirimycin inhibits the first enzyme (mannosidase I) in the pathway, and swainsonine inhibits the third glycosidase in the pathway (mannosidase II).[254] Studies with these latter inhibitors of glycan processing indicate that complex glycans are not necessary for secretion. Indeed, glycoproteins are secreted normally in sycamore cell suspensions and accumulate to normal levels when Golgi processing is inhibited.[255] The lack of a role for complex glycans in transport is supported by studies of a mutant of *A. thaliana* that is blocked in the conversion of high mannose to complex glycans due to an absence of the first glycosyltransferase enzyme in the pathway, *N*-acetylglucosaminyl transferase.[256] The mutant plants develop normally under several environmental conditions. Only further work will clarify whether this particular Golgi-mediated modification is critical for a small subset of glycoproteins elicited by a specific biotic or abiotic stress.[256]

Although N-linked glycans do not play a role in targeting per se, they may play a fundamental role in promoting correct protein folding and, as a consequence, enhance protein stability (e.g., via protection against proteolysis). Treatment of suspension-cultured plant cells with tunicamycin leads to a cessation of protein glycosylation and a decline in the accumulation of newly synthesized proteins in the culture medium and cell wall;[255-258] this is due primarily to an effect upon protein stability and not on protein synthesis.[259] Elimination of the glycosylation sites of phaseolin (a vacuolar storage protein in bean seeds) results in an increased susceptibility to proteolytic cleavage and decreased stability of the protein in transgenic tobacco seed.[260] A general effect of glycosylation on protein stability is indicated by similar studies of other seed vacuolar proteins (e.g., PHA and barley lectin).[82,251] The role of high-mannose glycans in the folding and assembly of soybean lectin was examined by determining the ability of different oligosaccharides to interfere with the reconstitution of the lectin from denatured subunit polypeptides.[261] The combined use of spectroscopy and size analysis by gel filtration revealed that both folding and assembly of denatured subunits were completely prevented in the presence of $Man_9GlcNAc_2Asn$ (M9-Asn). In contrast, the oligosaccharide GM9-Asn ($Glc_{1-3}Man_9GlcNAc_2Asn$) interfered with polypeptide assembly. These results suggest that the branch Man α1-2 Man α1-2 Man linked to the 3 position of the β-mannosyl residue of the high-mannose chains (Figure 13) functions in the folding of the subunit polypeptides, whereas assembly of folded subunits is associated with other branches.[261]

Glycosylation of proteins destined for secretion or the cell wall is very important in some plant species (e.g., carrot). The correctly glycosylated 52- to 54-kDa secreted protein is thought to play a crucial role in carrot somatic embryogenesis.[262] In tunicamycin-treated carrot cells, nonglycosylated cell-wall β-fructosidase is degraded in the endomembrane system immediately after synthesis.[259] The glycan heterogeneity characteristic of extracellular plant proteins may be important for the interaction of those proteins with the various components of the cell wall and cell surface (extracellular matrix).[239] Transport efficiency may also be affected by glycosylation. In animal cells, transport of the vesicular stomatitis virus G protein to the cell surface is completely abolished when the protein's glycosylation sites are eliminated by site-directed mutagenesis. Conversely, the creation of additional glycosylation sites in the protein (at certain positions) promotes transport, as a consequence of having a positive effect on protein folding.[60]

As discussed in the next section, newly synthesized proteins often associate with a variety of ER-resident molecular chaperones and folding enzymes to ensure that transport out of the ER lumen is limited to properly folded and assembled proteins. In animal cells, newly synthesized influenza virus hemagglutinin (HA) and vesicular stomatitis virus G protein associate transiently during their folding with calnexin, a membrane-bound ER chaperone. This association is dependent on the composition of the glycoprotein's N-linked glycans.[263] For example, inhibitors of N-linked glycosylation and of the enzymes involved in glucose trimming (glucosidases I and II) prevent an association of the nascent proteins with calnexin, whereas inhibitors of mannose trimming do not. Proteins with monoglucosylated glycans are the most likely binding species. Thus, a model has emerged that envisions the ER as containing a unique folding and quality control machinery in which calnexin acts as one chaperone that binds proteins with partially trimmed carbohydrate chains. Glucose trimming may provide a mechanism by which resident ER proteins can monitor the folding and maturation status of newly synthesized glycoproteins. The detailed configuration of the N-linked oligosaccharides may reflect the degree of folding, the state of oligomerization, duration in the ER, and so on. ER degradation may also be connected to trimming and calnexin association.[263] For example, the conditions that prevent glycoproteins from binding to calnexin also increase the degradation rate of some proteins.[264] While intimately involved in folding and oligomer assembly of glycoproteins, the calnexin system may have evolved mainly to ensure quality control. Although all newly synthesized glycoproteins so far tested bind transiently to calnexin, only a subgroup show folding and secretion defects in the presence of glucosidase inhibitors that inhibit this interaction.[265] When prevented from binding to calnexin then, many glycoproteins find alternative ways of folding and assembly in the ER lumen, the latter made possible by high concentrations of other ER-resident chaperones and redox enzymes.[263]

2. Folding and Oligomer Assembly

Among the numerous modifications that proteins of the secretory pathway undergo, one of the most important is the acquisition of a functional, three-dimensional structure and the assembly of polypeptide subunits into oligomers. With few exceptions, these processes take place in the lumen of the ER and in large part will determine a protein's stability and subsequent efficient transport (reviewed in References 4, 17, 59, 60, and 73). Many plant vacuolar proteins (e.g., storage proteins) are oligomers (commonly dimers, trimers, tetramers, or hexamers), most of which undergo complete oligomerization in the ER.[125,266] Most lectins and several enzymes destined for the vacuole or the cell surface are active in the ER, which is indicative that they have attained a functional three-dimensional conformation and are properly assembled.[49,245,267,268]

As mentioned in the previous section, transport of proteins out of the ER is subject to various rules of ER quality control and is often the rate-limiting step in transport along the secretory pathway.[16] Correct folding and assembly of proteins is necessary for their efficient exit from the ER and subsequent transport.[59] Thus, "sorting" of proteins at this step appears to be conformation specific, and failure to acquire the correct conformation can result in ER retention and/or degradation (see later discussion). The efficiency of protein folding *in vivo* is due, in part, to several resident proteins of the ER that play a role in assisting in both the folding and oligomer assembly of nascent proteins. Folding of most proteins likely begins before synthesis/translocation of the entire protein has been completed, as soon as a protein domain emerges into the ER lumen (reviewed in Reference 4). In fact, it appears that a nascent polypeptide chain is accosted on the luminal side of the ER membrane by a "welcoming" committee of enzymes and factors that facilitate and accelerate folding in several ways.[59,60,73,123,269–271] Perhaps the best characterized of these ER "helper" proteins (termed "molecular chaperones") is BiP (in mammalian cells also known as the glucose-regulated protein, GRP78).[131] BiP is related to the cytosolic heat-shock proteins (of the Hsp70 family), but is a soluble protein that is targeted to the ER by a cleavable signal peptide; a carboxy-terminal (H/K) DEL sequence mediates its retention.[70,131] Although the main function of BiP is probably to promote protein assembly, it may also serve to prevent export of proteins from the ER[59] (see later discussion). The existence of an ER-localized BiP has been demonstrated recently in tomato, tobacco, bean (*Phaseolus vulgaris*), spinach, and in the endosperm of maize and the aleurone layer of barley.[17,59,135,136,139,272–274] Although the plant and mammalian BiPs share sequence homology with the members of the 70-kDa heat-shock protein family, they are not themselves heat inducible.[4,131,139,272]

In addition to BiP, several other ER-resident proteins potentially involved in facilitating folding and assembly of nascent proteins have been identified recently in plants. One such protein is PDI (see later discussion). Others are calnexin and endoplasmin; the latter is an abundant ER resident also referred to as glucose-regulated protein 94.[275–277] Putative homologs of the mammalian endoplasmin have been identified in bean and tobacco.[141,272] The tobacco protein can be detected immunologically within the ER, and internal microsequence data reveals considerable sequence conservation to the mammalian endoplasmin.[4] Barley and periwinkle cDNA clones encoding 94-kDa proteins are identical in sequence to the tobacco counterpart within the region of the microsequence, and exhibit 78% identity at the amino acid level; both contain a C-terminal ER-retention signal, KDEL. As mentioned in the previous section, calnexin is another type of molecular chaperone that interacts with many nascent membrane and soluble proteins of the secretory pathway (reviewed in Reference 278). It is unrelated to BiP and endoplasmin and is further distinguished in that it is an integral membrane protein. A cDNA clone encoding a protein related to calnexin has been isolated from *A. thaliana*.[279] The plant calnexin cDNA encodes a protein of ~60 kDa, and shows an overall sequence identity of 48% to dog calnexin. The proteins from both organisms are ER associated and contain a large luminal domain followed by a single potential membrane-spanning domain near the C-terminus; a small C-terminal domain (that can be serine phosphorylated *in vitro*) is exposed to the cytoplasm (see Reference 279 and references therein). In pea plants, a calnexin-like protein is detected in microsomal fractions of different vegetative tissues (e.g., leaves, shoots, stems) and at different developmental stages (e.g., etiolated buds and light-grown leaves). Its constitutive expression is indicative of a fundamental function within many types of plant cells. A search is now underway to carry out structure-function studies focusing on the conserved sequences of the plant calnexin.[279]

As mentioned previously, the main function of BiP (and some of the other molecular chaperones) is probably to promote protein assembly; it also may serve to prevent export of misfolded proteins from the ER.[59] Synthesis of BiP is induced when abnormal proteins accumulate in the ER,[280,281] and it preferentially associates with such proteins until they either fold correctly or are degraded. Thus, by specifically retaining abnormally folded and incompletely assembled proteins, BiP contributes to the selectivity of transport.[59] Abnormal products retained in the ER by BiP include mutant and misfolded forms of proteins (e.g., influenza HA),[257] chimeric proteins, and some unglycosylated forms of proteins (e.g., invertase).[282–284] In an *in vitro* translation/translocation system, it binds to unoxidized, but not to mature, disulfide-bonded prolactin.[282] A role in promoting protein folding per se is suggested perhaps by helping to prevent the aggregation of folding intermediates. More specifically, proteins such as BiP may act as reversible detergents, binding to hydrophobic surfaces, but intermittently utilizing ATP to change their conformation to a nonbinding state.[285] In this manner, unfolded proteins would be maintained in solution without sequestering them permanently in a nonfunctional complex; the proteins could thus avoid aggregation or precipitation but still be able to achieve their final tertiary and quaternary structures.[59]

In addition to the ER, other organelles of the eukaryotic cell are supplied with their own set of "helper proteins." Cytosolic, mitochondrial, and chloroplast proteins belonging to the Hsp70 and Hsp60 families may function in a manner similar to BiP; like BiP, they are also ATPases that are characterized by ATP-independent binding of substrate and subsequent release in a step that is dependent on ATP hydrolysis (reviewed in References 4 and 278). Chaperones of the Hsp60 and Hsp70 families cooperate with each other, as well as with other chaperones, in the binding and release of unfolded proteins. During biosynthesis of the vesicular stomatitis virus G glycoprotein and the class I heavy chain IgG, there are interactions with both calnexin and BiP.[263,286] Thus, cooperation between chaperones may be a key factor in facilitating the folding of nascent proteins within the ER.[278] As intimated, secretory proteins may interact with certain molecular chaperones such as BiP and calnexin before they achieve their mature, correctly assembled, state.[282] In mammalian cells, BiP binds transiently in the ER to a portion of IgG heavy chains until assembly with the appropriate light chain occurs.[287,288] If no light chains are available, then the heavy chains remain associated with BiP until they are degraded. In a temperature-sensitive BiP mutant of yeast, BiP is essential for viability; import of proteins into the ER ceases within minutes of warming to the permissive temperature.[289] Thus, BiP interacts with a variety of nascent proteins and (in yeast cells, at least) may be required for some of them to complete their translocation into the ER,[59] a process aided by additional proteins (e.g., SEC61) that appear to be components of the translocon, a delicate structure closely apposed to polypeptides that are moving through the membrane (see earlier discussion).[93,113,115]

The ability of the tobacco BiP homolog to correct a temperature-sensitive BiP (*kar2*) defect in *S. cerevisiae*[139] provides strong evidence that the functional role of plant BiP in the ER is similar to that of the yeast and mammalian BiPs. The level of tobacco BiP is elevated in tissues of germinating seedlings and in organs containing specialized secretory tissues such as the anthers and stamens, in which the constituent cells likely have a large flux of protein through the secretory pathway.[139] ER isolated from aleurone layer cells of the barley grain contains a BiP cognate whose synthesis is upregulated by GA_3 and down-regulated by ABA.[273] In maize endosperm, the abundance of BiP and its mRNA increases dramatically following tunicamycin treatment (as occurs in mammalian cells). The antibiotic increases the amount of misfolded protein in the ER as a consequence of its inhibitory effect on protein glycosylation; thus, the requirement for BiP is increased. The association of BiP with polypeptides synthesized in the presence of tunicamycin (e.g., the bean seed storage protein phaseolin) can be reversed *in vitro* by the addition of ATP.[272,290] Presumably, the permanent misfolding of unglycosylated polypeptides causes continuous binding, release, and rebinding to BiP, lowering the concentration of unbound chaperone in the ER.[4] A similar increase in BiP and its mRNA occurs in mammalian and yeast cells in response to tunicamycin. Interestingly, yeast cells can monitor the concentration of free BiP in the ER and adjust the level of transcription of the BiP gene accordingly,[291] the latter mediated via a *cis*-acting element of the promoter (the 'unfolded protein response element'). This domain is similar in sequence to those located in the upstream regions of the mammalian BiP and *grp94* genes

accumulates in transgenic tobacco seeds to a much lower amount than its unmodified counterpart. However, the "hi-met" protein is transport competent and appears to be degraded in the vacuoles (protein bodies) and not in the ER.[312] In contrast, trimerization of phaseolin subunits is required for transport of phaseolin out of the ER in *Xenopus* oocytes;[229] monomeric phaseolin is retained in a pre-Golgi compartment until trimers are formed. A mutant of phaseolin generated by deleting a domain required for trimer formation remains as a monomer in *Xenopus* oocytes and is unable to reach the Golgi complex. Transient expression of the wild type and mutant phaseolin polypeptides in tobacco leaf protoplasts[290] indicate a similar recognition of assembly status in plant cells. The phaseolin mutant occurs in ATP-dependent association with BiP, strongly suggesting an ER location; there, the mutant polypeptides are rapidly degraded. Thus, the general features of a quality control system within the ER are conserved between plant and animal kingdoms. The extent of conservation of specific features will await greater understanding of the processes of ER retention, protein folding, and assembly and degradation of unfolded or defective proteins.[4]

3. Formation of Disulfide Bonds

Many proteins transported along the secretory pathway (e.g., seed storage proteins and extracellular proteins) have disulfide bonds that stabilize their tertiary and quaternary structures. The formation of these bonds is not random but is rapidly catalyzed in nascent chains by PDI (Figure 15) (reviewed in References 269 and 313). This enzyme interacts with unfolded proteins in the ER and catalyzes thiol oxidation and disulfide exchange reactions.[129] Depending on the nature of the polypeptide substrate and the imposed redox potential, PDI promotes disulfide formation, isomerization, or reduction. Mammalian PDI is a dimer of identical subunits, each of which contains two nearly identical putative active-site regions bearing close resemblance to the active site of the thioredoxins.[314] In mammalian cells, PDI has several distinct roles. Besides existing as a free monomer, it is also an essential subunit (the β-subunit) of prolyl-4-hydroxylase (an enzyme that catalyzes the modification of prolyl residues); PDI also binds to the Asn-X-Ser/Thr acceptor sequence for N-linked glycosylation and is an important component of oligosaccharide transferase, an enzyme that is essential for glycan transfer (i.e., as the glycosylation-site binding protein).[130] However, involvement of this enzyme during N-linked glycosylation in the ER is not demonstrated, and depletion of PDI from microsomes does not affect their capacity to elaborate oligosaccharides onto nascent polypeptides.[315] PDI has also been identified as a component of the microsomal triglyceride transfer protein complex.[316] In addition, the vertebrate enzyme shows substantial similarity in its amino acid sequence (including the two active sites) with rat phosphatidylinositol-specific phospholipase C (PI-PLC),[317] and contains two regions with sequence similarity to hormone-binding domains of the human estrogen receptor.[318] Mammalian PDI can be crosslinked to nascent immunoglobulin chains *in vivo*,[319] which may be indicative of a role in correct protein folding. However, PDI does not determine the polypeptide folding pathway, but rather facilitates formation of the correct set of disulfide bonds by promoting rapid reshuffling of incorrect disulfide pairings[269] (Figure 15). Also, in the presence of PDI, formation of the correct disulfide bonds is possible at higher oxidizing conditions compared with the spontaneous reaction.[320]

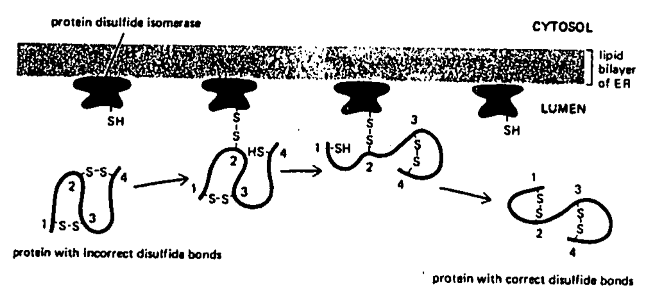


FIGURE 15. Action of protein disulfide isomerase. In the ER lumen, the enzyme protein disulfide isomerase acts repetitively to cleave intrachain S-S bonds until their arrangement has achieved the lowest overall free energy, at which point the protein is folded correctly. In this way, the enzyme facilitates the folding of the newly synthesized proteins that enter the ER. (From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D., *Molecular Biology of the Cell*, 2nd ed., Garland Publishing, New York, 1989. With permission from Garland Publishing.)

PDI has been detected in wheat, alfalfa, soybean, and tobacco, and some of the properties of these plant enzymes have been determined.[138,321-323] The wheat enzymes (from endosperm and aleurone layer cells) have molecular masses of ~130 to 150 kDa on gel filtration and a monomeric molecular mass of 57 to 60 kDa, suggesting that the native enzyme exists as a dimer.[322,323] The deduced amino acid sequence of alfalfa PDI contains the two thioredoxin-like active sites characteristic of vertebrate and yeast PDIs[324] and a C-terminal ER retention signal, KDEL.[128,138] Recently, immunogold labeling has localized PDI to the ER in soybean root nodules.[321] Whether the enzyme plays multiple roles in plant cells remains to be determined, although its importance for correct folding and disulfide bond formation has been established. All forms of PDI are soluble, or only loosely associated with the ER membrane, and are released by incubation at high pH. PDI-depleted microsomes can still import γ-gliadin (a nonglycosylated wheat storage protein), but the protein cannot achieve its correct disulfide-bonded state; reconstitution of microsomes in the presence of purified PDI restores this function.[315] Synthesis of the mRNA for alfalfa PDI is induced by tunicamycin, as is BiP mRNA; however, the protein induced by this treatment has a slightly lower molecular mass than in the absence of the drug.[138] The amount of PDI increases during seed development, reaching a maximum at 17 and 40 d of development in soybean seed and wheat endosperm, respectively.[129,321] In wheat grain, this corresponds to the period of endosperm development at which the rate of synthesis of disulfide-bonded storage proteins is maximal. PDI is also detected following germination and appears to be an indicator of seed viability. For example, PDI activity in microsome-enriched fractions from aleurone layers of normally germinating seeds is significantly higher than in their abnormally germinating counterparts and becomes more diminished in the latter as germination proceeds.[323]

In mammalian cells, ER proteins other than PDI also contain domains with homology to thioredoxin (e.g., ERp61 and ERp72, which are induced at the onset of immunoglobulin secretion in murine B cells).[144] In fact, it appears that the ER houses an extended range of enzymes that may use thioredoxin-like domains containing dithiol/disulfide active sites to carry out various functions in the co- and posttranslational modification of secretory proteins. It remains to be determined whether these proteins have additional roles in protein assembly that do not involve redox activity and whether they are residents of other compartments of the cell in addition to the ER.[269]

4. Amino Acid Modifications: Hydroxylation and Isoprenylation

Proteins that are secreted from plant cells such as extensins and other HRGPs of the cell wall[28] may also undergo modification to specific amino acid residues *en route* to the cell surface. For example, the biosynthesis of these proteins involves the hydroxylation of prolyl residues, followed by the glycosylation of many of the hydroxyprolyl residues.[325] These reactions are

carried out by 4-prolyl hydroxylase, an ER-associated enzyme and a glycosyl transferase that is associated with the Golgi complex.[212,326] It is not yet known whether the interesting protein homologies observed between 4-prolyl hydroxylase, PDI, and glycosylation site-binding protein of mammalian cells[130] are also present in plant proteins.[17]

The extensive posttranslational modifications to HRGPs (including hydroxylation, glycosylation, and crosslinking) are of importance to the role of these proteins in cell wall self-assembly and cell extension. Hydroxylation of proline residues may depend on multiple sequence-specific prolyl hydroxylases, rather than on a single conformation-dependent enzyme (reviewed in Reference 327). Furthermore, there are rules governing this modification; for example, Lys-Pro, Tyr-Pro, and Phe-Pro are not hydroxylated, whereas Pro-Val is consistently modified.

As mentioned previously, some proteins of the secretory pathway undergo isoprenylation, a modification that ultimately plays a role in plant growth, signal transduction, and membrane biogenesis (e.g., intracellular trafficking of membrane vesicles). Here, proteins are linked through a cysteinyl thioester bond to either a farnesyl or a geranylgeranyl moiety (reviewed in Reference 328). For some proteins (e.g., the small ras-like GTP-binding proteins), this modification is required for their assembly into membranes (including those of the ER, Golgi complex, and vesicles) and, hence, for their biological activity. In suspension-cultured cells of tobacco, there are at least two distinct prenyl:protein transferases. One utilizes farnesyl pyrophosphate and preferentially modifies a substrate protein with a CAIM carboxy-terminus (farnesyl:protein transferase). The other utilizes geranylgeranyl pyrophosphate and preferentially modifies a substrate protein that has a carboxy-terminus with the sequence CAIL (geranylgeranyl:protein transferase type I).[329]

5. Proteolytic Processing

Proteins transported along the secretory pathway often undergo posttranslational proteolytic processing, including seed storage proteins, defense-related proteins (destined for secretion or vacuolar transport), and various vacuolar enzymes (e.g., α-mannosidase and thiol proteases). For seed storage proteins, this may occur during their transit to the vacuole (i.e., along the secretory pathway), but more commonly occurs upon reaching their final destination (i.e., within the vacuole).[330] In some cases, the polypeptide is simply cleaved into two or three smaller polypeptides, all of which remain together in the oligomer; in others, a domain is lost (e.g., from the amino-terminus, carboxy-terminus, or middle of the polypeptide). The degree of processing is also variable; it may be complete or only partial, with some polypeptides remaining unprocessed (Table 6).[330] In the vacuole/protein body of the developing soybean seed, glycinin (a legumin-type storage protein) occurs as a 12S hexamer in which each of the six subunits is composed of an acidic and a basic polypeptide joined by a single disulfide bond.[331] Proglycinin subunits are assembled into 9S trimers in the ER, and, following their intracellular transport through the secretory system, arrive at the vacuole in that form. Within the protein body/vacuole, a specific cleavage occurs, yielding the acidic and basic polypeptides and the trimers assemble into hexamers[125,266] (Figure 16).[332] Assembly of monomers into trimers and then into hexamers occurs *in vitro*; this system has been used to evaluate the effects of protein modification on oligomer formation.[333] Interestingly, the assembly of glycinin trimers into hexamers requires the proteolytic cleavage of proglycinin into its basic and acidic subunits. For this reason, cleavage is considered a potential regulatory step in the pathway leading to glycinin deposition. The processing of glycinin is effected by a one-point cleavage that occurs at the carboxyl side of the asparaginyl residue located at the junction of the subunits, the modification being carried out by a cysteine endoprotease present in the vacuoles/protein bodies of soybean cotyledons.[334–336] Proteolysis is largely attributable to the conformational accessibility of the enzyme to the asparaginyl residue. Putative vacuolar processing enzymes are also involved in the cleavage of other classes of storage proteins, including the 2S albumins of *Arabidopisis*, *Brassica*, and other

TABLE 6
Types of Proteolytic Processing of Vacuolar Proteins

No proteolytic processing[a]	N-terminal domain lost
Phytohemagglutin	Napin
Soybean agglutinin	Sulfur-rich protein of brazil nut
β Subunit of soybean β-conglycinin	Tomato proteinase inhibitor I
Tomato proteinase inhibitor II	Potato proteinase inhibitor I
Pea vicilin	Sweet potato sporamin
Phaseolin	α and α′ subunits of soybean β-conglycinin
Patatin	
C-terminal domain lost	**Internal cleavage with or without loss of a domain**
Soybean glycinin	12S globulins
Wheat germ agglutinin	Sulfur-rich protein of brazil nut
Barley lectin	Napin
	Castor bean ricin
Cleavage and religation	Castor bean *Ricinus communis* agglutinin
Concanavalin A	Pea lectin
	Pea vicilin

[a] Except signal peptide.

From Chrispeels, M. J. and Tague, B. W., in *Recent Advances in Development and Germination of Seeds*, Taylorson, R. B., Ed., Plenum Press, New York, 1990. With permission from Plenum Press.

seeds;[337] the thiol protease of barley aleurone layer cells (aleurain) is also processed by a specific endoprotease located within the vacuole.[338]

Of all seed storage proteins studied, the 2S albumins[339] undergo some of the most extensive posttranslational processing. Maturation is a multistep process that follows cleavage of the signal peptide, and involves the removal of three peptides: an amino-terminal, an internal, and a carboxy-terminal fragment.[340,341] In the mature protein, subunits of 9 and 4 kDa remain that are linked by disulfide bonds. Posttranslational cleavages occur on the C-terminal sides of asparagine residues 35 and 74, which are conserved among the 2S albumin precursors in seeds of different plant species.[342] The two Asn residues are located in the hydrophilic regions of the precursor, suggesting the vacuolar processing enzyme can recognize exposed Asn residues on the surface of the protein and effect cleavage on the C-terminal side of these residues.[342]

As noted earlier, certain defense-related proteins, including chitinases, endo β-glucanases, and α-amylase inhibitors also undergo proteolytic processing. For the α-amylase inhibitor of bean seeds, cleavage upon arrival at the vacuole/protein body serves to activate the protein, possibly by removing a conformational constraint on the precursor.[343]

Some of the proteases involved in the postgerminative mobilization of seed storage proteins have been identified (see References 344 through 352); however, their transport to the vacuole has not been characterized.

E. Transport of Proteins Through the ER and Golgi Complex

All of the transport steps from the ER to the Golgi complex and transport through the Golgi itself are common to most proteins of the biosynthetic/secretory pathway and occur via a signal-independent (bulk-flow) mechanism (see earlier discussion). Proteins destined for locations beyond the ER are transported in small membranous vesicles that bud from the smooth regions of the ER membrane. These so-called 'transition' vesicles are directed to the Golgi apparatus which is composed of flattened, membrane-enclosed saccules. In animal cells, the Golgi apparatus comprises three distinct compartments, each con-

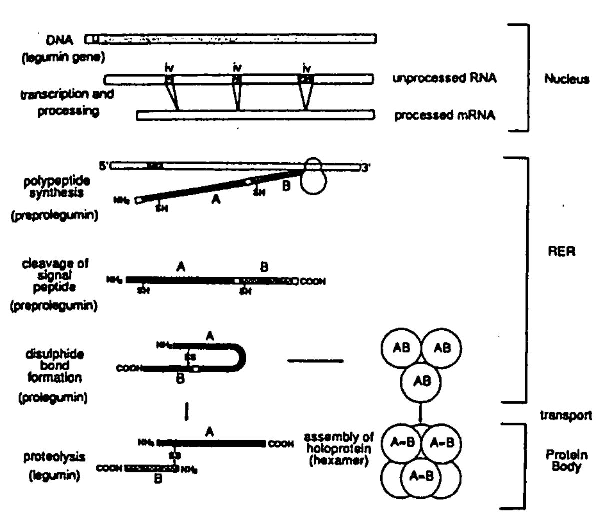


FIGURE 16. A representation of the synthesis and processing of the legumin (glycinin) in soybean cotyledons. Following transcription, the mRNA encoding glycinin is transported from the nucleus to the rough endoplasmic reticulum (RER), where it is translated. The mRNAs for the acidic (A) and basic (B) subunits are joined by codons for a 4-amino-acid linker sequence on the primary translation product (preprolegumin). The mRNA also contains a code for the signal peptide at the amino-terminal end and a pentapeptide at the carboxy-terminal end. The primary product (prolegumin) is processed to yield the A and B subunits joined by the linker sequence and disulfide bonds (S-S); the signal peptide and the pentapeptide are cleaved off. In the final step of processing the mature protein is formed as the linker sequence is removed, and the acidic and basic subunits are joined only by the disulfide bonding. The mature acidic subunit contains 278 amino acids (approximately 40 kDa) and the basic subunit 180 amino acids (approximately 20 kDa). Assembly of the subunits within the protein bodies yields the mature hexameric legumin holoprotein. Based on Krochko and Bewley.[332] Reprinted with permission of Longman Group UK (Gale Research, Inc.). See also Dickinson et al.[286] for details and Bednarek and Raikhel[83] for a general review. (From Bewley, J. D. and Black, M., *Seeds. Physiology of Development and Germination*, 2nd ed., Plenum Press, New York, 1994. With permission from Plenum Press.)

taining a characteristic array of protein-modifying enzymes, the *cis*, *medial*, and *trans* cisternae, plus the tubular TGN.[353] Movement through the Golgi compartments is mediated by the budding and fusion of transport vesicles. Much of the present work in animal cells and yeast is focused upon the mechanisms of protein movement through the Golgi stacks and how proteins are sorted according to destination, then packaged for delivery.[15] In both systems, signal-dependent sorting of proteins away from bulk flow (i.e., to the lysosome/vacuole or to secretory vesicles) occurs in the TGN, a complex tubular reticulum in which proteins are segregated into different transport vesicles to be dispatched to their final destinations.[56,354] Thus, in this compartment protein traffic signals are recognized by specific receptors that sort proteins according to their proper destination. For example, soluble lysosomal enzymes are diverted to lysosomes by a mechanism involving a protein sorting signal (a modification of a N-linked carbohydrate chain) and a membrane-bound receptor that recognizes the specific carbohydrate modification (reviewed in Reference 85). Likewise, within cells involved in regulated secretion, selected proteins undergo sorting and dense packing into special vesicles in the TGN.[355]

1. Vesicles Involved in Transport

In mammalian cells, the vesicles that transport proteins from the TGN to other compartments (via sorting-dependent mechanisms) are covered with a protein coat consisting of clathrin[356] and a subset of characteristic polypeptides.[357] Small clathrin-coated vesicles function in both receptor-mediated endocytosis as well as in the receptor-mediated transport of proteins to the lysosome (reviewed in References 16 and 358 through 360). The endocytic and lysosomal transport pathways meet in the TGN. The participation of mannose-6-phosphate receptors in both transport pathways is mediated by independent domains on the receptor molecule.[361–363] A cytoplasmic tail domain on the mannose-6-phosphate receptor directs its clustering in Golgi-localized clathrin-coated buds, mediated by a number of 'adaptor' molecules. From these buds, vesicles form, carrying the receptor and its ligand for transport to lysosomes and eventual recycling of the receptor to the Golgi complex. A second domain directs entry of the receptor into clathrin-coated pits on the plasma membrane mediated by a different set of adaptor molecules. The two domains on the receptor protein function independently and do not compete for the same set of adaptor molecules.[362,363]

As intimated above, coated pits concentrate selected vesicle cargo, usually transmembrane receptors bound to a soluble ligand, while excluding resident membrane proteins. The cytoplasmic

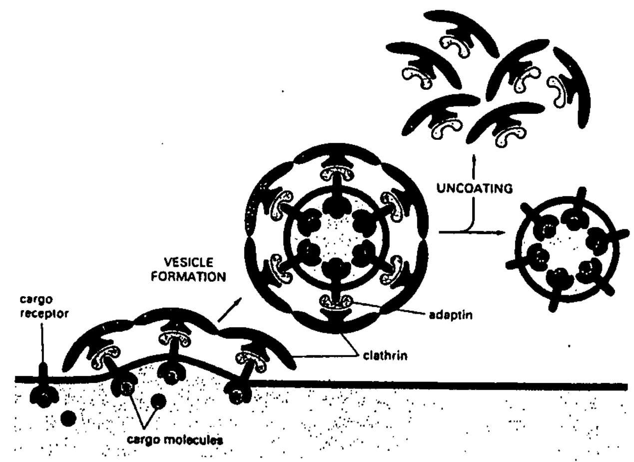


FIGURE 17. Selective transport mediated by clathrin-coated vesicles. The adaptins bind both clathrin triskelions and cargo receptors. (From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D., *Molecular Biology of the Cell*, 3rd ed., Garland Publishing, New York, 1994. With permission from Garland Publishing.)

face of the coated pit is composed of a layer of clathrin-associated proteins (adaptins) bound to the membrane, surrounded by an attached lattice of clathrin subunits thought to be assembled by the recruitment of soluble forms of clathrin and adaptins to the membrane (reviewed in Reference 364). Coated pits progress to form vesicles; once released from their membrane attachment, mature coated vesicles are rapidly uncoated, releasing clathrin and adaptins for another round of coated pit and vesicle formation (Figure 17). The uncoated vesicle is now a fusion-competent vesicular intermediate. An outstanding question is whether clathrin is involved in the export of constitutively secreted proteins and normal membrane proteins. One view is that the contribution of clathrin to intracellular traffic is quantitative rather than absolute; clathrin assembly concentrates and sorts receptors, facilitating their transport from one membrane to another.[359] It performs this function in both the import and export pathway, enhancing transport provided by bulk flow in either direction. Localized control of clathrin assembly through diverse molecular signals from receptors and coated-pit components allows its universal function to be applied to specific intracellular targeting.

The major structural component of the clathrin-coated vesicle is the clathrin triskelion, a three-legged structure comprised of three clathrin heavy chains (180 kDa) and three clathrin light chains (30 to 40 kDa), each bound to a heavy chain near the vertex of the triskelion.[365,366] Clathrin heavy chains may play a structural role in vesicle formation, effecting a change in membrane shape required to produce a curved vesicle.[367] The clathrin light chains, on the other hand, likely play a regulatory role in vesicle consumption by regulating the equilibrium between clathrin coat assembly and disassembly (reviewed in References 364 and 368). For example, disassembly of the clathrin lattice and release from the vesicle surface requires clathrin light chain, and is stimulated by the enzymatic action of Hsc70 (uncoating ATPase), a member of the heat-shock protein family.[369]

The adaptins of coated pits and vesicles are thought to mediate clathrin binding to membranes and play a role in the selection of vesicle cargo.[370,371] Two major adaptin complexes designated AP-1 and AP-2 (AP: assembly or adaptor protein) have been characterized biochemically and localized; AP-2 is associated with clathrin on the plasma membrane,[357] whereas AP-1 is found exclusively in association with clathrin in the TGN.[372] These adaptin complexes are thought to link the clathrin lattice with the membrane by binding both to clathrin and to transmembrane proteins in the coated vesicle membrane.[373] Their position within the 20 nm space between the vesicle surface and clathrin lattice is consistent with this proposed role; likewise, adaptin complexes are able to bind *in vitro* to the cytoplasmic domains of transmembrane receptors that are packaged into clathrin-coated vesicles.[359,363] For example, the cytoplasmic domain of the LDL (low-density lipoprotein) receptor binds the adaptin complex AP-2. Thus, they are likely to initially bind selected trafficking proteins in coated pits. When the lattice-coated membrane pinches off, all of these proteins are incorporated into free carrier vesicles.[374] Further elucidation of clathrin-adaptin interactions may come from reconstituted clathrin-coated vesicles and/or a genetic approach (e.g., in yeast).[364]

A vesicle with only a partial clathrin coat is involved in the dense packing of proteins secreted via a regulated (and sorting-dependent) mechanism.[375] In the formation of hormone-containing secretory granules, clathrin-coated pits and vesicles remove membrane from the site of a condensing secretory bud that will eventually pinch off from the TGN to form a secretory granule; the prohormone may be initially concentrated by a receptor-mediated process (reviewed in Reference 359). After granule acidification and prohormone processing, the receptors may be removed by clathrin, resulting in granule condensation to form the mature secretory granule.[56] This scheme again postulates that the role of clathrin and the associated adaptor complexes is to concentrate and sort receptors. It is now evident that the sorting of proteins destined for regulated secretion occurs upon exit from the TGN, where they are directed into vesicles distinct from those involved in the transport of secretory proteins of the constitutive pathway; this sorting also occurs faithfully in a cell-free system.[355] A small nonclathrin-coated vesicle is involved in bulk-flow transport; in contrast to the clathrin-coated vesicles, these carriers are found at all levels of the Golgi stack and contain protein at its prevailing (bulk) concentration in the parental Golgi cisternae (see later discussion).[376,377]

In the cells of higher plants, the structural relationships between the ER and Golgi apparatus and between intraorganellar transport pathways are much less clear.[17,29] However, there is evidence that some features of Golgi structure and vesicular transport are similar to those of animal cells. For example, ultrastructural analysis of the individual Golgi stacks in suspension-cultured sycamore cells identified three morphologically distinct cisternae.[235] In plant cells, physical connections between the *trans*-Golgi and the partially coated reticulum (an organelle consisting of tubular membranes bearing clathrin-like coats over parts of their cytoplasmic surface) are indicative that the latter resembles the TGN of animal cells.[378] Clathrin-coated vesicles are involved in endocytosis in plant cells;[358,379] they may also participate in the sorting-dependent transport of proteins to the plant cell vacuole. For example, precursors of vacuolar seed storage proteins and lectins are found in clathrin-coated vesicles isolated from developing pea cotyledons.[380–382]

Research on plant clathrin-coated vesicles has been impeded in large part by the difficulty of obtaining sufficient quantities of intact, pure, and undegraded vesicles with which to work.[2,383] The heavy and light chains of plant clathrin are 10 to 15 kDa larger than their animal counterparts (Figure 18).[384–386] Most of the adaptor polypeptides (as well as receptors) remain to be identified; however, there is now evidence for a β-type adaptin (i.e., a plant equivalent of the β-adaptin component of the complex AP-2 of mammalian cells).[387] A major 28-kDa polypeptide in clathrin-coated vesicle fractions of pea cotyledons, initially thought to be an adaptin, has now been identified as a contaminant (phytoferritin).[383] This latter study has underscored the limitations of using certain plant tissues as a source for clathrin-coated vesicle isolation (e.g., embryonic tissues such as cotyledons that store large amounts of ferritin and photosynthetic tissues in which postmicrosomal fractions are contaminated with the chloroplastic enzyme, ribulose bisphosphate carboxlase-oxygenase).

As mentioned previously, uncoating of clathrin-coated vesicles is a prerequisite to vesicle fusion with the target membrane. Recently, a cytosolic uncoating ATPase (an enzyme that dissociates clathrin from clathrin-coated vesicles in the presence of ATP) has been isolated from developing pea cotyledons using chromatography on ATP-agarose.[388] The properties of the plant enzyme are markedly similar to those of the uncoating ATPase of mammals and yeast. Thus, despite the dissimilarities in the molecular weights of the clathrin components of vesicles from their respective sources, uncoating appears to be achieved by a common mechanism.[388] Characterization of additional coat components of plant vesicles and sequence analysis of identified coat proteins will allow a more detailed investigation of the uncoating process; the mechanism of uncoating may be elucidated by uncoating experiments carried out *in vitro*, in which clathrin baskets lack specific coat components.[388]

2. Mechanisms of Vesicle Budding, Targeting, and Fusion: Cytosolic Factors, Receptors, and GTP-Binding Proteins

Insight into the mechanisms of vesicle budding, targeting, and fusion has come from the use of a variety of experimental approaches in both animal cells and yeast (reviewed in References 389 through 396). Semi-intact cells made permeable to macromolecules have been used to study ER to Golgi traffic.[397,398] In this system, there is efficient reconstitution of intracellular transport, provided ATP and a cytosolic fraction are added.[397] *In vitro* cell-free systems have been developed for yeast.[399] The secretory pathway in yeast has been partially defined by a series of temperature-sensitive mutant strains that are blocked at various stages in protein transport.[400–403] Such mutants were used to show that a number of complementation groups (and thus gene products) are required for the transport of secretory proteins from their site of synthesis to the cell surface; several are required for ER to Golgi transport alone. Additional gene products are necessary for intra-Golgi transport and for localization of yeast hydrolases to the lysosome-like vacuole. To inves-

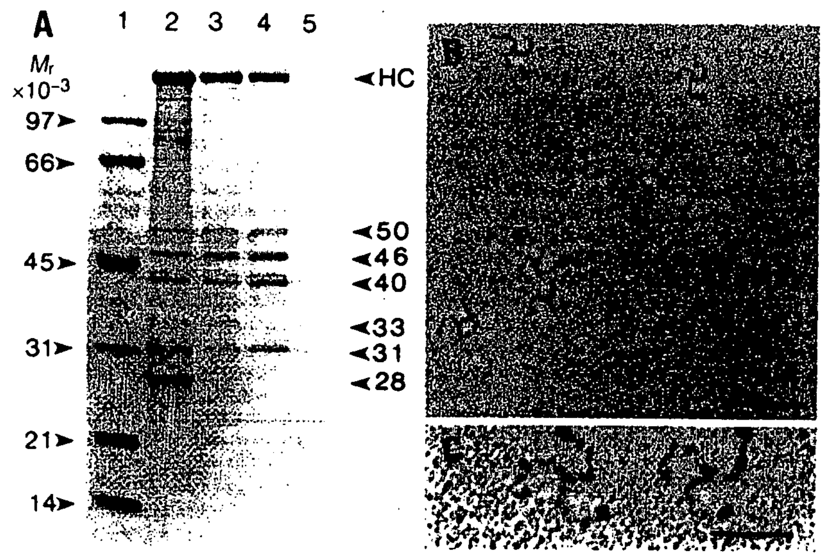


FIGURE 18. Clathrin triskelion of pea coated vesicles (B and C) and protein composition (A). In (A), the protein composition of the triskelions was examined by SDS-PAGE and stained with Coomassie blue. Lane 1, relative molecular mass markers; lane 2, total coated vesicle proteins; lane 3, coat proteins dissociated by buffer with 2 *M* urea; lane 4, proteins in triskelion fraction; lane 5, proteins in a second A_{280} fraction. Arrowheads indicate position of HC (190), 50-, 46-, 40-, 33-, 31-, and 28-kDa polypeptides. (From Lin, H. B., Harley, S. M., Butler, J. M., and Beevers, L., *J. Cell Sci.*, 103, 1127, 1992. With permission from the Company of Biologists, Ltd. Courtesy of L. Beevers.)

tigate the function of these gene products, researchers are cloning these genes and characterizing the proteins they encode (reviewed in Reference 404). Secretory pathway mutants can be complemented by unique expression plasmids that contain segments of wild-type yeast DNA. Open reading frames are identified (by DNA sequence analysis), gene fusions to β-galactosidase are constructed and gene product-specific antibodies are generated.[16] In this manner, some of the proteins required specifically for ER to Golgi transport and for intra-Golgi transport have been identified and characterized (Figure 19).

Vesicular transport between cisternae of the Golgi stack has been demonstrated *in vitro*.[16] Transport of the vesicular stomatitis virus G protein from the *cis*- to the *medial*-compartments of the Golgi complex was shown to require a crude cytosolic fraction, ATP, and proteins isolated from the surface of the Golgi membranes (Figure 19).[405–407] A number of cytosolic factors (proteins) are involved in collecting proteins for vesicular transport, as well as mediating vesicle budding, transport, and fusion of the transport vesicle with the appropriate target organelle. Remarkably, the cytosol from yeast[408] and plants[409] can substitute for animal cell cytosolic extracts in promoting transport in the Golgi stack and in forming coated vesicles. This interchangeability is indicative that the transport machinery is extremely similar, even in detail, in all eukaryotes. More recently, techniques that were developed with rat liver were used to isolate a population of vesicles derived from ER in a cell-free system

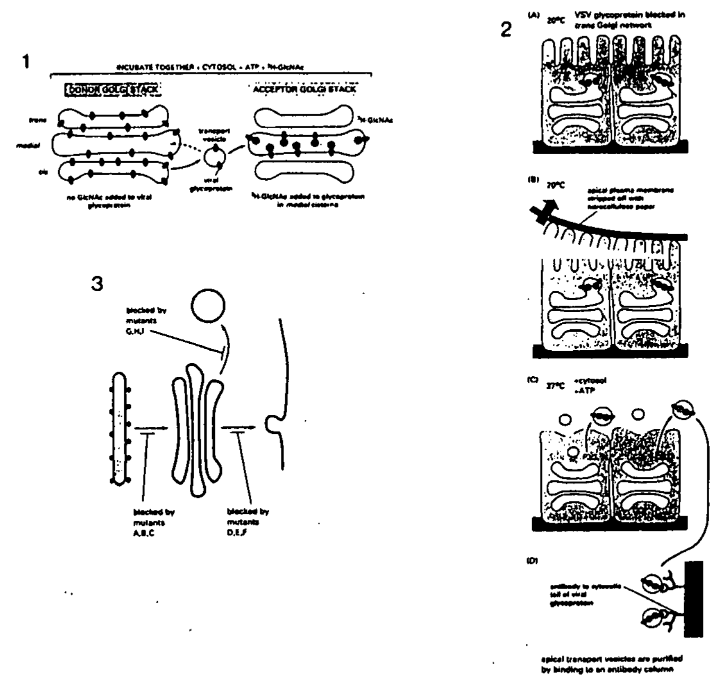


FIGURE 19. Approaches to the study of ER to Golgi and intra-Golgi transport in yeast and mammals. (1) Cell-free systems for studying the components and mechanism of vesicular transport. Reconstituted vesicular transport in cell-free systems was first achieved for the Golgi stack. When Golgi stacks are isolated from cells and incubated with cytosol and with ATP as a source of energy, transport vesicles bud from their rims and appear to transport proteins between cisternae. By following the progressive processing of the oligosaccharides on a glycoprotein as it moves from one Golgi compartment to the next, it is possible to follow the process of vesicular transport. To follow the transport, two distinct populations of Golgi stacks are incubated together. The "donor" population is isolated from mutant cells that lack the enzyme *N*-acetylglucosamine (GlcNAc) transferase I and that have been infected with a virus; because of the mutation, the major viral glycoprotein fails to be modified with GlcNAc in the Golgi apparatus of the mutant cells. The "acceptor" Golgi stacks are isolated from uninfected wild-type cells, and thus contain a good copy of GlcNAc transferase I, but lack the viral glycoprotein. In the mixture of Golgi stacks, the viral glycoprotein acquires GlcNAc, indicating that it must have been transported between the Golgi stacks — presumably by vesicles that bud from the *cis* compartment of the donor Golgi and fuse with the *medial* compartment of the acceptor Golgi. This transport-dependent glycosylation is monitored by measuring the transfer of ^{3}H-GlcNAc from UDP-^{3}H-GlcNAc to the viral glycoprotein. Transport occurs only when ATP and cytosol are added. By fractionating the cytosol, a number of specific cytosolic proteins have been identified that are required for the budding and fusion of transport vesicles. (2) Semi-intact cells. Vesicular transport can also be studied in cells whose plasma membrane has been permeabilized to allow small molecules and macromolecules to leave and enter the cell freely. Permeabilization is achieved by physical rupture or treatment with bacterial toxins that punch large holes in the plasma membrane. Such semi-intact cells are particularly useful for studying the transport from extended membrane systems that become extensively fragmented during conventional homogenization procedures, such as the ER and the *trans* Golgi network. Semi-intact cells have been used to isolate transport vesicles that mediate transport from the *trans*-Golgi network to the apical plasma membrane. (3) Genetic approaches for studying vesicular transport. Genetic studies of mutant yeast cells defective for secretion at high temperature have identified more than 25 genes that are involved in the secretory pathway. Many of the mutant genes encode temperature-sensitive proteins; these function normally at 25°C, but when the mutant cells are shifted to 35°C, some of them fail to transport proteins from the ER to the Golgi apparatus, others from one Golgi cisterna to another, and still others from the Golgi apparatus to the vacuole or to the plasma membrane. (From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D., *Molecular Biology of the Cell*, 3rd ed., Garland Publishing, New York, 1994. With permission from Garland Publishing.)

from plants that likely represent an intermediate compartment in membrane transfer between the ER and Golgi complex.[410]

The finding that proteins can transfer from one Golgi stack to another in a unidirectional fashion indicates that vesicles have a built-in targeting capacity; specifically, each type of vesicle has its own "address marker" that ensures that inappropriate transport does not occur. The target membrane must in turn have a complementary receptor that recognizes the address and allows docking of the vesicle.[15,16] This type of specificity is involved in transport from the ER to the Golgi and in intra-Golgi transport. Proteins from mammalian and yeast intra-Golgi transport vesicles and from Golgi membranes that appear to fulfill these functions have been characterized; however, the identity of some of these tags remains elusive. The working concept for such targeting is that during vesicle transit through the cytosol, the configuration of the polypeptides on the vesicle surface mediates vesicle docking and release of contents only when the correct membrane is encountered. Indeed, recent evidence indicates that vesicle docking involves the formation of a complex that includes proteins from the cytoplasm as well as proteins that are embedded in the vesicle and target membranes (reviewed in Reference 389). After reaching the TGN, proteins are sorted according to their correct final destination.

Vesicles responsible for intra-Golgi transport contain nonclathrin coat proteins (coatomer proteins or COPs), in addition to other associated proteins, including a small GTP-binding protein, ADP ribosylation factor (ARF). A cell-free assay developed for intra-Golgi transport in mammalian cells led to the identification and purification of a cytoplasmic protein needed for vesicles to fuse with their target membranes, called NSF (for NEM [*N*-ethylmaleimide]-sensitive fusion protein).[411] NSF was later found to be functionally interchangeable with SEC19 of yeast (reviewed in Reference 412). NSF carries two repeat motifs for an ATP-binding site, and ATP binding is critical for NSF-induced fusion.[393] Further work led to the identification of soluble NSF attachment proteins (SNAPs) — cytosolic proteins that associate with NSF to allow membrane association.[393] α-SNAP, one of the three SNAP proteins associated with NSF in mammalian cells, also has a yeast counterpart (i.e., SEC17).[412] The yeast protein SEC4, a ras-like GTP-binding protein, is crucial at a later stage in the secretory pathway, required for fusion with the cell membrane;[413] mammalian equivalents of the yeast SEC4 protein have also been identified (referred to as *rab*, for *ras* genes/proteins from rat brain; see later discussion). Thus, at least three protein families — NSF, SNAPs, and Rabs — are all cytoplasmic proteins that transiently associate with membranes.[389] Figure 20 presents a working model for vesicular transport between compartments of the Golgi complex that is envisioned as a GTP hydrolysis-driven cycle. ARF is myristylated at its amino-terminus that, in conjunction with its GTP-bound state, appears to regulate ARF association with membranes.[414,415] More specifically, ARF-GTP causes the binding of other coat components to the Golgi membranes, leading to the budding of coated vesicles. In this model, target membranes possess a GTPase activity that causes the hydrolysis of the ARF-bound GTP; thus, encounter with the target membrane destabilizes the coat (i.e., causes vesicle uncoating). This would recycle ARF and the coatomers to the cytosol and lead to the fusion of the uncoated vesicle with the target membrane.[393] A recently identified interaction between ARF and phospholipase D (PLD) activity links together the dynamics of coat formation/removal with phospholipid metabolism (reviewed in Reference 390). To add to this model, other groups of proteins important in the vesicle fusion process function as receptors for some of the soluble factors. These receptor proteins (referred to as SNAREs [SNAP receptors]) are thought to be permanently embedded in the membranes to act as markers and binding sites for the cytoplasmic proteins. These integral proteins, then, are of two varieties — those that are integrated into the vesicle membrane (v-SNARE) and those that reside in the target or acceptor membrane (t-SNARE).

An NSF-SNAP affinity matrix has been used to isolate putative SNAREs from bovine brain membrane extracts.[416] Purified recombinant NSF and SNAPs were mixed with detergent-solubilized bovine brain microsomes in the presence of ATPγS, and any complexes that formed were

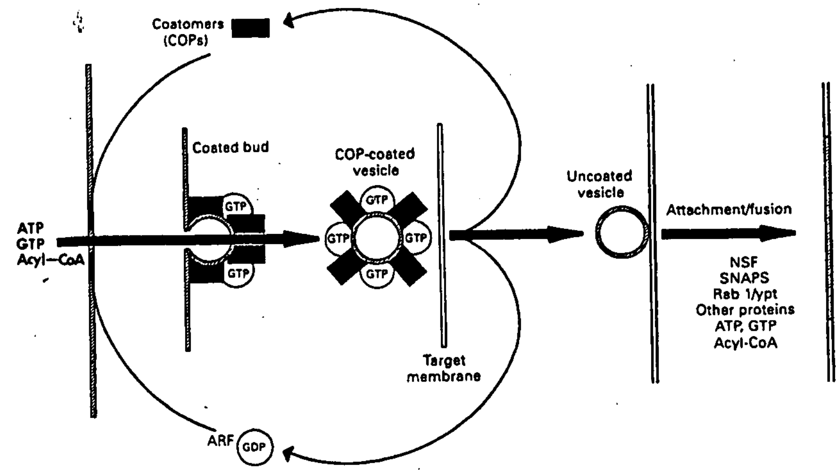


FIGURE 20. Working model for vesicular transport between Golgi compartments. The shaded membrane at left represents a portion of a Golgi cisternal membrane giving rise to a nonclathrin COP-coated vesicle that attaches to its target membrane (unshaded), the envelope of the next Golgi compartment in the secretory pathway. Specific targeting of the coated vesicle is proposed to be initiated by an unknown component in the recipient membranes. The vesicle is then uncoated and the general NSF-dependent fusion pathway is triggered, together with attachment of the now uncoated vesicle. Components of the coated vesicle include, but may not be limited to, coatomers (a complex of COPs, filled squares) and ADP-ribosylation factor (ARF), which can exist in both GDP- and GTP-bound forms (circles). The proposed cycle for coatomers and ARF has not been directly demonstrated. (From Rothman, J. E. and Orci, L., *Nature*, 355, 409, 1992. With permission from Nature and Macmillan Magazines Ltd.)

immunoisolated. Subsequently, the proteins that had bound in an ATP-dependent fashion were eluted using Mg-ATP. The eluate contained SNAPs and also a number of putative SNAREs, including syntaxins A and B and synaptobrevin (or VAMP); these proteins were previously identified as key membrane components of synaptic vesicles and the synaptic plasma membrane. Their localization was indicative of a role in synaptic-vesicle-plasma-membrane fusion. More direct evidence of a functional role in targeting/fusion came with the demonstration that NSF and SNAP interact to form a 20S complex. A model has now emerged in which the NSF-SNAP complex functions as a bridge between receptor proteins for SNAPs (SNAREs) in vesicular and target membranes of a fusion reaction.[395] The v-SNARE (the vesicle address tag) interacts with a target-membrane SNARE (t-SNARE), specifying targeting and leading to fusion by the SNARE interaction with SNAPs and NSF (Figure 21).[417] In brain synapses, syntaxins act as SNAREs in the target membranes (t-SNAREs) (i.e., the presynaptic cell membrane), whereas the synaptic vesicle protein synaptobrevin (VAMP) acts as the respective receptor in the vesicle membrane (v-SNARE).[416,418] Alternatively, the NSF-SNAP complex may act on a preexisting complex formed by syntaxin, synaptobrevin, and a third protein associated with the presynaptic membrane, SNAP-25 (synapse-associated protein).[418] Interestingly, the syntaxins (t-SNAREs) share sequence homology with several yeast proteins (SSO1, SED5, and PEP12) that are involved in membrane transport (Table 7).[412,419] The recent findings suggest a general model for vesicle targeting and fusion in

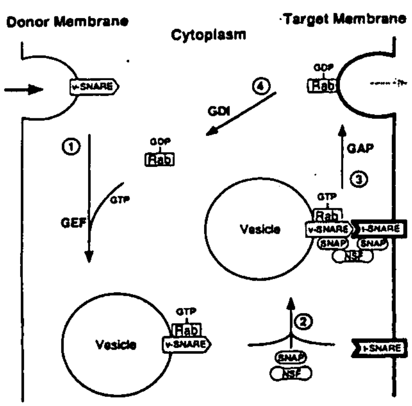


FIGURE 21. The SNARE hypothesis for vesicle fusion with the target membrane. According to this hypothesis, the v-SNARE protein located on the vesicles interacts with the t-SNARE found on the target membrane in the presence of NSF, SNAPs, small GTP-binding proteins, and other uncharacterized components. This interaction leads to the eventual fusion of the transport vesicles with the target membrane. In step 1, the small GTP-binding protein Rab is recruited to the vesicle in the presence of GEF and GTP. The recognition and interaction between v-SNARE and t-SNARE mediated by NSF and SNAP bring the vesicle to the target membrane (step 2). Hydrolysis of GTP by Rab with the help of GAP triggers the fusion of two distinct lipid bilayers (step 3). The GDP-bound Rab is released by GDI from the membrane to the cytosol for recycling (step 4). Fusion of the vesicle to the target membrane also releases the cargo of the vesicle for secretion into or outside of the membrane compartments (shaded arrow). The solid arrow indicates the formation of vesicles. (From Verma, D. P. S., Cheon, C.-I., and Hong, Z., *Plant Physiol.*, 106, 1, 1994. With permission from the American Society of Plant Physiologists.)

which the soluble protein NSF and the SNAPs (or their yeast counterparts SEC18 and SEC17) are common to fusion complexes throughout the cell, whereas the integral membrane proteins VAMP and syntaxin represent protein families whose members are located on different types of vesicles and take part in specific kinds of fusion.[389]

On a cautionary note, the model may be oversimplified because it was originally built on work on neurons in which release of neurotransmitter is regulated and not constitutive. However, the general SNAP-SNARE concept as a mechanism for directing fusion events in Golgi trafficking appears to operate in both constitutive and regulated secretion. It is likely that the regulated secretion mechanisms of neurons and other specialized secretory cells of animals use components of the general secretory machinery, but in addition, contain specialized components to make it regulated and able to respond to a specific stimulus. One

TABLE 7
Traffic Patterns: Mammalian Proteins and Their Yeast Counterparts

Protein	Alternate names	Location	Yeast counterparts
NSF		Cytoplasm	SEC18
SNAPs (α,β,γ)		Cytoplasm	SEC17(a)
VAMP	Synaptobrevin	Vesicle membrane	BET1 and SLY2 (ER to Golgi) SNC1 and 2 (Golgi to plasma membrane)
Syntaxin	HPC-1	Plasma (target) membrane	SED5 (ER to Golgi) SSO1 and 2 (Golgi to plasma membrane) PEP12 (Golgi to vacuole)
SNAP-25		Cytoplasm and plasma membrane	

From Barinaga, M., *Science*, 260, 486, 1993. With permission from the American Association for the Advancement for Science.

possible candidate for such a regulator is the vesicle protein synaptotagmin, which binds calcium and is not found in cells capable of only constitutive secretion (e.g., yeast).[420] The search is now on for other members of the v- and t-SNARE families — particularly the latter, because these proteins likely play very important roles in defining the fusion site on transport vesicles and in establishing and maintaining organelle integrity.[392]

Genetic and biochemical studies in yeast and mammalian cells have also led to significant advances in the identification of proteins that are involved in mediating ER to Golgi transport as well as elucidating the role of GTP-binding proteins as regulatory elements in the secretory pathway. By combining the power of yeast genetics with a permealized-cell transport system, Lian and Ferro-Novick[421] have implicated two vesicle integral membrane proteins, BOS1 and SEC22 (both v-SNAREs), that are required for fusion of vesicles involved in ER to Golgi transport with target Golgi membranes as well as the specificity of the membrane fusion event. Antibodies to BOS1 inhibit ER to Golgi transport *in vitro*, without affecting vesicle budding. The putative t-SNARE that interacts with BOS1 or SEC22 is SED5.[422] The *sed5* gene was isolated as a suppressor of the *erd2* mutation. As mentioned earlier, ERD2 is the KDEL receptor and is required for stability of the Golgi complex. When ERD2 is overexpressed, Golgi proteins appear to redistribute to the ER, and the Golgi complex disappears.[166] Depletion of the ERD2 protein, on the other hand, causes membranes to accumulate.[422] Evidence implicating SED5 as a t-SNARE that is required for consumption of transport vesicles is indicated by the finding that the temperature-sensitive *sed5* yeast mutant exhibits an accumulation of ER-derived vesicles; overexpression of SED5 also leads to vesicle accumulation, but these likely represent vesiculated Golgi membranes. Thus, SED5, like ERD2, may be required to maintain Golgi integrity.[392] Other membrane proteins implicated in ER-Golgi transport are BET1 and YPT1; the latter is a small Rab-type GTPase that is incorporated in ER-Golgi transport vesicles.[421] BET1 may represent a t-SNARE necessary for recycling vesicles in the retrograde pathway from the Golgi back to the ER.[392]

GTP-binding proteins have been implicated as key regulatory elements in the secretory pathway (reviewed in References 391 and 423 through 436). Notably, in the last decade, more than 30 different GTP-binding proteins of the Ras superfamily have been implicated in the regulation of membrane traffic at virtually every stage of the exocytic and endocytic pathways (Table 8).[391,417] In very general terms, GTP binding and hydrolysis controls switching between two different pro-

TABLE 8
Localization and Possible Function of Small GTP-Binding Proteins in Different Membrane Compartments

Protein	Localization	Possible function
Ras	Plasma membrane	Signal transduction
Ran	Nucleus	Nuclear protein import, DNA synthesis
Rac	?	Phagocytosis, membrane ruffling
Rho	Golgi, post-Golgi vesicles	Actin cytoskeleton
Arf	Golgi	Regulate budding from the ER and fusion at the Golgi stacks, endosomes, and nuclear vesicles
Sar	ER	Vesicular budding from the ER
Rab		
Rab1a	?	ER-Golgi transport
Rab1b	ER, Golgi	ER-*cis*-Golgi transport
Rab2	ER-Golgi intermediate compartment	ER-Golgi transport
Rab3a	Synaptic vesicles, chromaffin granules	Regulated exocytosis
Rab3b	Mainly in cytosol	Regulated exocytosis
Rab3c	Synaptic vesicles	Regulated exocytosis
Rab4	Early endosomes	Early endosome-plasma membrane recycling pathway
Rab5	Early endosomes, plasma membrane	Plasma membrane-early endosome transport, fusion of early endosomes
Rab6	TGN, post-Golgi transport vesicles	Budding from TGN
Rab7	Late endosomes	Transport in endocytic pathway
Rab8	Post-Golgi basolateral secretory vesicles	Golgi-plasma membrane transport
Rab9	Late endosomes, TGN	Late endosome-TGN transport
Rab11	TGN, secretory granules, synaptic vesicles	?
Rab12	Golgi	?
Rab13	Tight junction	Polarized transport, assembly of tight junction
Rab17	Basolateral plasma membrane	Transcellular transport
Rab22	Plasma membrane, endosomes	Endocytic pathway
Rab24	ER, *cis*-Golgi, late endosomes	Endocytic pathway

Note: In Budding Yeast; Ypt1: ER to Golgi, intra-Golgi transport; Sec4: Golgi to Plasma Membrane transport; Ypt7: Endocytosis.

From Verma, D. P. S., Cheon, C.-I., and Hong, Z., *Plant Physiol.*, 106, 1, 1994. With permission from the American Society of Plant Physiologists.

tein conformations; the switch serves to propagate and amplify regulatory signals. A subset of these proteins (G proteins) may mediate the vectorial transport of individual vesicles, specifying the direction of vesicular transport between appropriate cellular compartments.[404,423,427] The low molecular weight monomeric GTP-binding proteins fall into two subgroups. The first subgroup includes proteins of the ARF/Sar family that are involved in the formation of carrier vesicles from the donor organelle and in the preparation of the vesicles for fusion with the acceptor organelle (reviewed in Reference 394). The other family includes SEC4, YPT, and Rab proteins that control distinct vesicular transport events (Table 8).

A clear indication of the importance of GTPases is the inhibitory action of nonhydrolyzable GTP analogs such as GTPγS on, for example, intra-Golgi transport[428] and budding of secretory vesicles from the TGN.[355] GTPγS acts by locking the GTPase in an activated state.[429] For example, ARF is the cytosolic protein that confers sensitivity of a cell-free Golgi transport assay to GTPγS.[430] A ras-like GTP protein, Sar1, is required for ER to Golgi transport. Like ARF, the association of the Sar1 protein with ER membranes is GTP-dependent. A facilitatory role is played by a 70-kDa integral membrane glycoprotein (SEC12, a resident of the ER), which promotes the association of Sar1 with the ER and thereby promotes protein transport from the ER to the Golgi complex. SEC12 has been identified as a GDP/GTP exchange protein (see later discussion).[431] Another participant is SEC23, a GTPase-activating protein (GAP) specific for the Sar1 protein; in carrying out its role, a complex is formed with another protein required for vesicle budding, SEC24.[432] The model that emerges (Figure 22) then is a process in which Sar1, with the aid of SEC12, binds GTP, associates with the ER, and then (helped by SEC23/SEC24) hydrolyzes GTP to catalyze the formation of a transport vesicle. Although not shown in Figure 22, coatomers are not only important in intra-Golgi transport, but are also required for ER to Golgi transport (see Reference 394 and references therein).

As noted above, the second subgroup of GTP-binding proteins — SEC4, YPT, and Rab proteins — control distinct vesicular transport events (Table 8). Genetic evidence indicates that different members of the SEC4/YPT/Rab protein family have unique cellular functions and, for the most part, cannot replace each other. For example, overexpression of the *ypt1* gene will not compensate for deletion of the *sec4* gene and vice versa. The yeast *ypt1* gene encodes a ras-like protein that binds and hydrolyzes GTP; this protein resides in the Golgi complex and in ER-Golgi carrier vesicles and is required for transport from the ER to and through the Golgi complex.[421,426,433] For example, the yeast mutant strain *ypt1* accumulates vesicles and antibodies to the YPT1 protein block ER to Golgi transport *in vitro*. Interestingly, mammalian homologs of the YPT1 protein,[434–436] Rab1a and Rab1b, are 75 and 66% identical to the YPT1 protein, respectively.[436] Like YPT1 of yeast, Rab1 is essential for both ER to Golgi and intra-Golgi transport in mammalian cells *in vivo* and *in vitro*.[437,438]

The protein SEC4 plays an essential role at late stages of the yeast secretory pathway. Most of the SEC4 protein is attached to the cytoplasmic faces of both post-Golgi secretory vesicles and the plasma membrane; the remainder is soluble in the cytoplasm.[439] This diversity in location suggests that proteins of the SEC4/YPT/Rab family may undergo a cycle of localization as they fulfill their functions.[391,425,426] Newly synthesized SEC4 rapidly associates with secretory vesicles that go on to fuse with the plasma membrane. The plasma membrane-bound pool can recycle onto a new round of vesicles as they accumulate in a secretory mutant blocked in exocytosis; this recycling

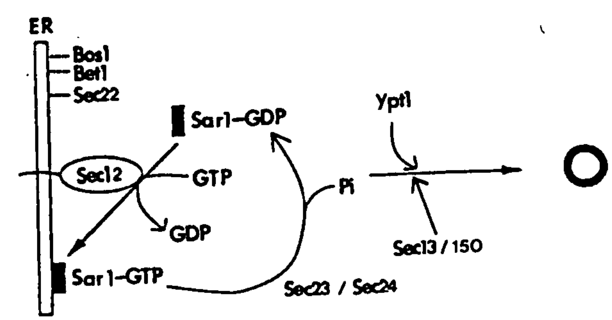


FIGURE 22. Some proteins involved in ER to Golgi transport in yeast. See text for details. ER-derived vesicles contain Sec22, Bos1, and Ypt1. (From Takizawa, P. A. and Malhotra, V., *Cell*, 75, 593, 1993. With permission from Cell Press.)

pathway may utilize a soluble intermediate. The SEC4 protein may act primarily on the vesicle's surface to transduce an intracellular signal needed to regulate transport between the Golgi complex and the plasma membrane.[439] Another particle (SEC8/SEC15) is found in the cytosol and is peripherally associated with the plasma membrane, but not with secretory vesicles. It may function as a downstream effector of the SEC4 protein, serving to direct the fusion of vesicles with the plasma membrane; there is indirect evidence of a weak association of SEC4 with this complex.[413]

In plants, virtually no functional studies of GTP-binding proteins have been undertaken due, in part, to the lack of *in vitro* model systems and the difficulty of screening for mutants that are defective in the endomembrane transport system.[4] However, the recent identification of cDNA clones encoding proteins highly related to the yeast YPT family (e.g., in maize, rice, pea, soybean, and *Arabidopsis*) suggests that the mechanisms of vesicular transport in higher eukaryotes are conserved (reviewed in Reference 417).[440–446] In maize and *Arabidopsis*, the proteins predicted from these cDNAs show high similarity to other members of the Ras family, particularly in the regions involved in GTP/GDP binding, GTPase activity, and membrane binding.[443] Further experiments using reverse-genetic analysis with antisense constructs or dominant-interfering mutations, and cytological analyses using specific antibodies, may elucidate the precise roles of YPT proteins in vesicle trafficking and plant growth.[443]

Homologs of the ARF/Sar family of GTP-binding proteins have been identified recently in *Arabidopsis* and tomato.[447–450] A clone (*A.t. Rab6*) encoding a small GTP-binding protein homologous to Rab6 of mammals has been isolated from a cDNA library of *Arabidopsis* leaf tissue.[446] The mammalian Rab6 protein (homologous to the proteins Ryh1 in *Schizosaccharomyces pombe* and YPT6 in *S. cerevisiae*) is associated with the *medial*- and *trans*-Golgi cisternae as well as with the TGN.[451,452] Disruption of the *S. cerevisiae* homolog (YPT6) results in partial missorting of vacuolar proteins,[453] indicating its participation in late events of the secretory pathway related to vacuolar targeting. Functionally, the plant gene is able to complement the temperature sensitive phenotype of the *ypt6* null mutant in yeast.[446] Characterization of this gene and its product will be awaited with interest and will allow a more complete dissection of the machinery involved in soluble protein sorting at the TGN.[446]

Secretory mutants in *A. thaliana* are presently being sought.[330] Although specific defects in transport to the plant cell vacuole are the primary focus of these researchers, a more complete set of secretory mutants may help to identify specific proteins important in mediating the various transport steps of the biosynthetic/secretory pathway. It is also noteworthy that cells of the giant alga Chara have been permealized,[454,455] and it may be possible to use this type of system to study protein secretion.[2] Compounds that stimulate or inhibit secretory events (e.g., Ca^{2+} and GTPγS, respectively) can also be microinjected into actively growing plant cells,[456] and isolation of protoplasts derived from cells that are active in protein secretion (e.g., cereal aleurone layer cells) allows analyses at the single-cell level.[457] Use of these techniques in conjunction with a molecular genetics approach should help to clarify the role of GTPases in plants.[2,458]

As noted above, GTPases of the Rab/YPT1/SEC4 family go through a characteristic cycle of reactions that drive the transition between at least two distinct conformational states (Figure 23).[459] Release of GDP from the 'inactive' (GDP-bound) state allows binding of GTP.[429,460] It then returns to the inactive state after GTP hydrolysis, an irreversible reaction that renders the cycle unidirectional. The function of GTPases as molecular switches depends on the abilities of these distinct conformational states to interact with specific macromolecules. Superimposed upon this are various regulatory or accessory proteins that control the GTPase cycle, such as guanine nucleotide dissociation inhibitors (GDI) that inhibit GDP dissociation, guanine nucleotide exchange proteins (GEP) that stimulate GDP dissociation, and the GAPs that promote GTP hydrolysis (reviewed in References 461 and 462). Recent progress has been made on the identification of these various types of accessory proteins that are specific for members of the SEC4/YPT/Rab family. A GAP that acts on YPT6 has been cloned recently from yeast by screening colonies containing high copy number plasmids with genomic inserts for increased GTPase activation.[463] Also in yeast, *Dss4* (encoding a 17-kDa protein) facilitates SEC4 protein action *in vivo* by functioning as a GDP dissociative stimulator (i.e., a GEP).[464] Likewise, studies on a *rab1* mutant suggest a cycle in which the function of a Rab1-specific GEP, which mediates guanine nucleotide exchange, is critical for recruitment of Rab1 during vesicle budding and the formation of ER and Golgi carrier vesicles competent for fusion with downstream compartments.[462] GTP hydrolysis is likely to be critical also for a late Ca^{2+}-dependent vesicle targeting fusion step controlling the delivery of vesicles to Golgi compartments.[465] Finally, a protein that acts to inhibit the dissociation of GDP from a broad range of Rab proteins has been purified from bovine brain[461,466] and yeast.[467] These proteins (termed Rab-GDI) remove Rab proteins from membranes leading to the formation of a cytosolic complex of Rab with the inhibitor protein (reviewed in Reference 391). A multistep mechanism is involved in the membrane attachment of Rab into the donor membrane (following dissociation of the Rab-GDI complex), which is mediated by a guanine-nucleotide exchange factor.[468] Thus, these various accessory proteins may function together to control each step of a functional cycle that couples GTP binding and hydrolysis to membrane attachment and recycling (Figure 21).[391,417]

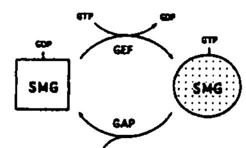


FIGURE 23. Small GTP-binding proteins as molecular switches. The switch is turned on when a small GTP-binding protein (SMG) binds to GTP with the help of GEF. GAP promotes GTP hydrolysis, which turns the switch off. (From Verma, D. P. S., Cheon, C.-I., and Hong, Z., *Plant Physiol.*, 106, 1, 1994. With permission from the American Society of Plant Physiologists.)

The current view also incorporates an interaction between Rab proteins with vesicle receptor proteins (Figure 21).[391] More specifically, receptors incorporated in the vesicle surface (v-SNAREs) may only be active for docking in the presence of the appropriate Rab protein in its GTP-bound form. Following vesicle fusion, hydrolysis of the GTP bound on the Rab protein could inactivate the v-SNARE on the target membrane. Regulation of SNAREs by Rabs could be mediated by a direct interaction, or there could be intermediates to the regulatory pathway.

F. Hormonal and Environmental Controls over Protein Secretion in Plants

The study of the effects of hormones and environmental (stress) conditions on protein transport along the endomembrane system has been hindered in large part because other "upstream" regulatory controls are often affected simultaneously (e.g., gene transcription, mRNA stability, protein synthesis, and protein folding).[4] One must also distinguish between direct and indirect effects of the altered physiological conditions on ER-mediated functions. In some cases, changes in ER function brought on by stress or hormonal changes affect the synthesis and/or transport of only a specific subset of proteins. An example is the effect of heat shock on the synthesis of seed storage proteins and their transport out of the ER.[469,470] In bean (*Phaseolus vulgaris*) cotyledons subjected to heat stress, the synthesis of phaseolin storage protein is reduced; conversely, synthesis of PHA, another major storage protein in the bean seed, is increased.[469] There is likewise an enhanced synthesis of storage proteins within developing soybean (*Glycine max*) seeds in response to heat shock.[470] In heat-shocked bean cotyledons, a large proportion of PHA accumulates in a microsomal fraction and is not transported to the Golgi complex, nor is it found in the vacuoles, the normal site of deposition of PHA. Electron-dense material accumulates within the ER cisternae in the heat-shocked cells, perhaps indicating the presence of protein aggregates. The elevated temperature may promote misfolding and aggregation of

PHA within the ER lumen, impairing its transport out of that compartment (i.e., to the Golgi complex, *en route* to the vacuole).[469] Interestingly, the heat shock is not accompanied by an increased synthesis of BiP — a protein expected to play a major role in retention of misfolded/misassembled proteins within the ER; however, other molecular chaperones playing a similar role may be induced (e.g., the low-molecular-weight Hsps).[4]

Studies on hormonal controls over protein synthesis and secretion have been most extensive on the aleurone layer of cereals (reviewed in Reference 49). The aleurone layer is comprised of small thick-walled cells that surround the starchy endosperm of the cereal grain (Figure 24). Following germination of the mature grain, these living cells secrete a large number of hydrolytic enzymes into the starchy endosperm.[49] Here, they hydrolyze and degrade the stored reserves, including polymers of the starch grains, protein bodies, and cell walls, thus providing nutrition for the enlarging and differentiating embryo. The products of this degradative process are thought to enter the embryo through uptake by the epithelial cells of the scutellum. These scutellar cells also secrete hydrolytic enzymes into the starchy endosperm following germination. The control of reserve mobilization in the cereal grain was established around 1960: the embryo produces gibberellin (GA), which induces the living cells on the periphery of the endosperm, the aleurone layer, to produce a number of enzymes. For some of these enzymes, the promotive effect of GA is on their synthesis only; for others, GA also stimulates their secretion (e.g., α-amylase, carboxypeptidase, ribonuclease, acid phosphatases).

The most well-studied enzyme, synthesized and secreted by the aleurone layer in response to GA and largely responsible for starch hydrolysis, is α-amylase, an enzyme that represents as much as 70% of the newly synthesized protein in GA_3-treated barley aleurone cells. A multitude of complex and interrelated events constitute, coordinate, and control the secretory response of the barley aleurone layer; superimposed upon the hormonal controls is a central role played by calcium, which is just beginning to be elucidated. α-Amylase is secreted from the aleurone cell along the constitutive pathway via the Golgi apparatus.[29] Following entry of the nascent protein into the lumen of the ER, the enzyme is posttranslationally modified in a number of ways. Two of these modifications (that may be related) are the folding of the protein into a specific and functional three-dimensional conformation and the addition of Ca^{2+}.[471] Barley α-amylase is a Ca^{2+}-containing metalloprotein whose activity and stability are dependent on the binding of at least one atom of Ca^{2+} per mole of α-amylase. Both the folding of α-amylase into a functional three-dimensional conformation and the addition of Ca^{2+} may be regulated by the ER-resident, BiP (see below). Ca^{2+} binding also leads to a marked change in the tertiary structure of the α-amylase molecule, as shown by changes in tryptophan fluorescence following the removal of bound Ca^{2+} from the protein.[34]

The dependence on Ca^{2+} for activity/stability of α-amylase has impeded elucidation of the precise role of Ca^{2+} on synthesis/secretion of the enzyme;[472] nonetheless, considerable progress has been made in the last 5 years. Less than 1% of the Ca^{2+} needed for the synthesis of active α-amylase is available within the aleurone cell.[473] Ca^{2+} binding to α-amylase is estimated to require a Ca^{2+} concentration of at least 10 μM within the ER lumen; thus, the synthesis and secretion of α-amylase by isolated aleurone layers is highly dependent upon the supply of external Ca^{2+} required at millimolar concentrations (e.g., 10 mM). Although synthesis and secretion of α-amylase is promoted by GA_3 in the presence of Ca^{2+}, this process is prevented by abscisic acid (ABA). Controls by GA and ABA with respect to induction or suppression of α-amylase synthesis and secretion by isolated aleurone layers appear to be exerted in part at the level of transcription. The 5′ upstream *cis*-acting sequences responsive to ABA and GA have been identified in α-amylase genes, and the *trans*-acting factors (DNA-binding proteins) potentially involved in regulating gene expression are being characterized.[474–478] However, hormonal controls over α-amylase synthesis and secretion exist at additional (posttranscriptional) levels; these controls in turn appear to be intimately related to the requirement for Ca^{2+}. These various controls (Figure 25) (reviewed in Reference 34) include the following:

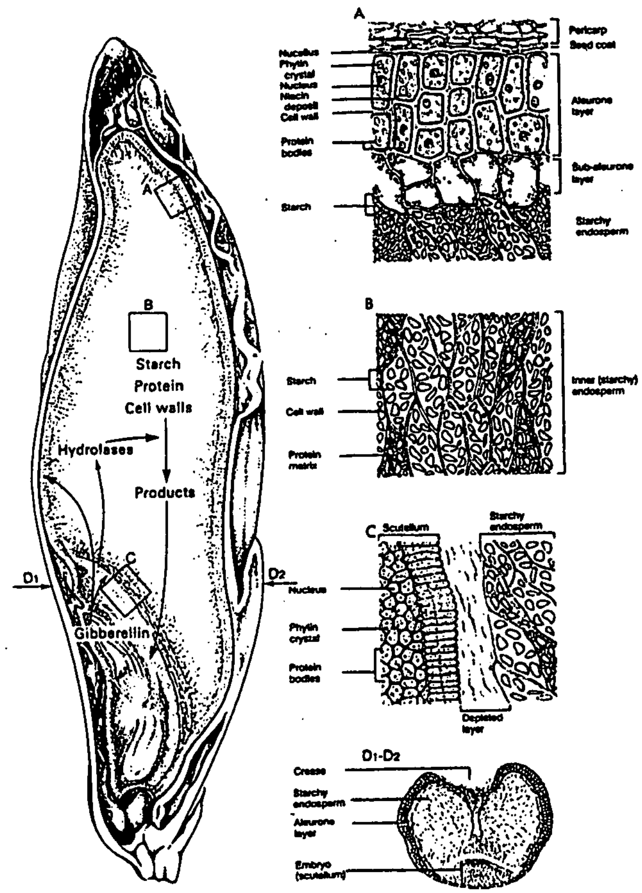


FIGURE 24. The structure of the barley grain. On the left is a longitudinal section of the whole grain. (A) The aleurone and sub-aleurone layers. (B) Endosperm cells. (C) The interface between the scutellum and the endosperm. (D 1–2) Transverse section through the grain where the embryo and endosperm overlap. (From Jones, R. L. and Jacobsen, J. V., *Intl. Rev. Cytol.*, 126, 49, 1991. With permission from Academic Press.)

1. GA and ABA control calcium homeostasis in aleurone cells. GA has a profound effect on the flux of Ca^{2+} into the aleurone cell, which is increased by approximately tenfold, possibly as a consequence of the hormone stimulating the opening of channels in the plasma membrane and/or preventing efflux of Ca^{2+} from the cytosol. The GA-

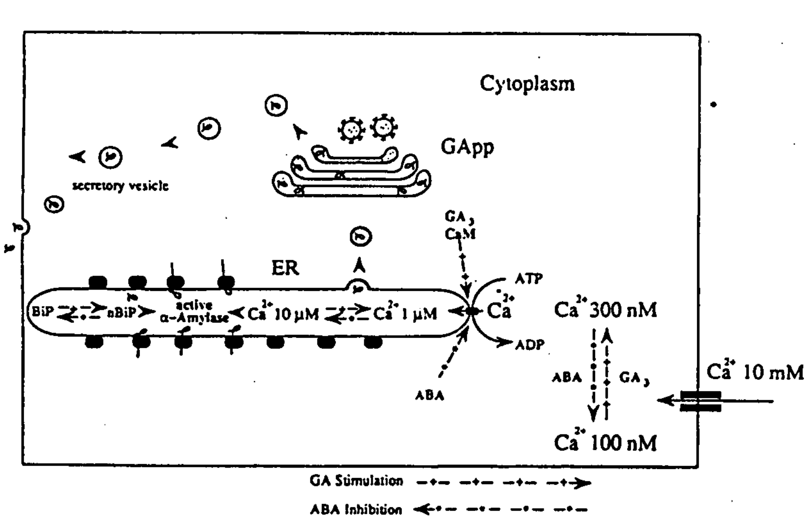


FIGURE 25. Model of the calcium-dependent events occurring during hormonal regulation of secretory activity in the barley aleurone cell. (From Jones, R. L., Gilroy, S., and Hillmer, S., *J. Exp. Bot.*, 44 (Suppl.), 207, 1993. With permission from Oxford University Press.)

induced change in cytosolic Ca^{2+} is most pronounced within the periphery of the cytosol and precedes the onset of α-amylase synthesis and secretion. ABA reversal of GA-induced α-amylase synthesis is preceded by a lowering of cytosolic Ca^{2+} levels.[479]

2. GA and ABA regulate calcium uptake into the ER. The two hormones appear to regulate Ca^{2+} uptake by the ER at several levels. First of all, the activity of an ATP-dependent Ca^{2+} transporter located on the ER membrane of barley aleurone cells is stimulated by GA and inhibited by ABA.[480] Increased cytosolic Ca^{2+} may play an important role in regulating the transport of Ca^{2+} into the lumen of the ER *in vivo*; the increase as a consequence of GA_3 treatment likely stimulates the activity of the Ca^{2+} pump. As a consequence, the ER can accumulate millimolar levels of calcium.[479,481] GA also increases calmodulin (CaM) levels in the aleurone cell, which presumably further activates the ER Ca^{2+} transporter.[482] The increased Ca^{2+} flux into the ER supports the synthesis of α-amylase (a calcium metalloprotein), and also activates and stabilizes the molecule.[34] The accumulation of Ca^{2+} in the ER may be facilitated further by the presence of Ca^{2+}-binding proteins within the ER lumen that also play roles as molecular chaperones. The amount of a BiP cognate within the ER lumen of barley aleurone cells is elevated by GA and reduced by ABA.[273] Thus, BiP may play an important role in allowing the ER to accumulate Ca^{2+}, in addition to aiding in the folding of the α-amylase molecule. These various controls would be expected to affect α-amylase transport and secretion because secretory proteins are generally not competent for transport out of the lumen of the ER until they have assumed the correct three-dimensional conformation.

3. Calcium and CaM may act as secondary messenger molecules mediating hormone-induced changes in ion channel activity. The activity of a slow vacuolar (SV) ion channel in the tonoplast of barley aleurone storage protein vacuoles is increased in protoplasts treated with GA_3, and decreased in the presence of ABA.[41] The opening of the channel is sensitive to cytosolic-free Ca^{2+} concentrations $[Ca^{2+}]_i$ between 600 n*M* and 100 μ*M*. Ca^{2+} may activate endogenous CaM that is tightly associated with the membrane; CaM also sensitizes the SV channel to $[Ca^{2+}]_i$. The function of this particular SV channel in barley aleurone cells is unknown; it may be involved in mobilization of stored K^+ from the aleurone cells to the embryo following germination. Nonetheless, these results indicate that the SV channel, like the plasma membrane and ER Ca^{2+}-ATPases, is a plant ion transporter regulated by Ca^{2+}-CaM. Because GA regulates both $[Ca^{2+}]_i$ and CaM levels in barley aleurone cells, Ca^{2+} and CaM are proposed to act as signal transduction elements mediating hormone-induced changes in ion channel activity.[41] The mechanism of the CaM effect on channel gating (e.g., via a kinase or phosphatase) is being investigated presently.[41]

In summary, by regulating the elements involved in the signal transduction pathway, phytohormones can regulate the synthesis and secretion of the α-amylase molecule.[34] Identification of GA and ABA receptors has not been achieved yet. However, recent studies have yielded convincing evidence that GA and ABA are perceived initially on the external face of the plasma membrane (see review by Allan and Trewavas[483]). For example, barley aleurone protoplasts respond to GA_3 and ABA applied extracellularly, but not to hormone microinjected into the cytosol of the cell;[484] membrane impermanent GAs have biological activity (i.e., are capable of eliciting α-amylase synthesis) in oat protoplasts, also indicative of a surface receptor for GA action.[485] Elucidation of the entire signal transduction pathway (e.g., possible regulation of G protein activity following hormonal perception) will be awaited with interest. Some of the events that GA triggers may be specific to α-amylase production and secretion. For example, in contrast to α-amylase, a fusion protein consisting of the inert marker molecule PAT linked to a signal peptide, is secreted efficiently when synthesized transiently in barley aleurone protoplasts treated with ABA.[486] Because secretion was equally efficient in ABA-treated vs. GA-treated protoplasts, ABA does not have an adverse effect on the general capacity of aleurone cells for secretion.

G. Signal-Mediated Sorting of Proteins to the Plant Cell Vacuole and Retention in ER-Derived Protein Bodies

As mentioned previously, the plant cell vacuole is a multifunctional compartment; its functional diversity arises in part from the strict spatial and temporal regulation of the genes encoding soluble and membrane proteins destined to reside there (reviewed in Reference 18). The central vacuole of mature plant cells occupies more than 80 to 90% of its total volume; one of its major roles is to regulate turgor, important for the mechanical stability of plants. Vacuoles of many cells are used as an intermediate storage compartment for ions, sugars, and amino acids; in these cell types there is a regular exchange of these compounds between the cytosol and vacuole. Carrier or channel proteins located within the vacuolar membrane (tonoplast) often mediate this exchange, in which the driving force for uphill transport can be a proton gradient generated by tonoplast-associated H^+-ATPases and H^+-pyrophosphatases.[18] Vacuoles of other plant cell types serve as storage depots, such as those for storage of defense proteins and allelochemicals, or for storage proteins.

1. *Mechanisms of Storage Protein Deposition in Seeds*

Transport and targeting of proteins to the plant cell vacuole have been investigated most extensively in seeds, particularly in relation to the synthesis and deposition of storage proteins (reviewed in References 12, 17, 83, and 487 through 492). As mentioned previously, within many seeds,

vacuoles perform a dual function — as temporary storage depots (during seed development) and as sites of macromolecular hydrolysis (during postgerminative seedling growth). The synthesis and deposition of seed storage proteins are subject to strict temporal and spatial regulation; they only occur in specific tissues/organs and only at specific times during seed development. Cell expansion and deposition of reserves (including storage proteins), follow the histodifferentiation stage of seed development, and occur largely in the absence of further cell divisions (reviewed in References 35 and 36). It is during this time that young storage parenchyma cells (particularly within the major reserve organs of seeds — the endosperm, cotyledons, or megagametophyte) accumulate a variety of proteins, including storage proteins, acid hydrolases, plant defense proteins, and other reserve materials or metabolites in storage compartments (organelles) called protein bodies.[330] In vegetative tissues, leaf and bark storage proteins accumulate in vacuoles that resemble the protein bodies of seeds (Figure 26).[23,493]

It is noteworthy to mention here that there are distinct pathways of protein body formation in seeds (Figure 26). One pathway that is characteristic of most dicotyledonous seeds involves the subdivision or fragmentation of a large central vacuole. In this mechanism of protein body formation, which also appears characteristic of the gymnosperm seeds studied so far,[494,495] storage proteins are deposited into subdividing vacuoles (protein bodies) via transport through the Golgi complex (e.g., rice globulins [glutelins], wheat prolamins, and legume globulins). The globulins of leguminous seeds are likely soluble in the ER and undergo subsequent transport to the vacuole (via the Golgi complex) as soluble components. However, after arrival at the vacuole, they likely precipitate due to the low pH of the vacuole (and possibly also as a result of proteolytic processing) (reviewed in Reference 489; see references therein). A second mechanism occurs in some cereals and involves direct formation of protein bodies from the rough ER (reviewed in References 488, 489, and 496 through 500). Here, storage proteins are synthesized on the rough ER and are deposited directly within the lumen of this organelle, where they remain and accumulate (e.g., prolamins of maize, rice, and sorghum). There is now evidence for yet another route (e.g., for some of the wheat prolamins), which begins by assembly of the proteins into protein bodies within the ER. However, these ER-derived protein bodies undergo subsequent internalization within vacuoles by a specific process analogous to autophagy.[4,488,489] Thus, the wheat prolamin storage proteins exhibit two pathways or routes of transport to protein bodies.

2. Biogenesis of ER-Derived Protein Bodies and Mechanisms of Protein Retention

In many cereals (e.g., maize, rice, and sorghum), the prolamins are retained within the lumen of the rough ER to form protein deposits that remain surrounded by the ER membrane (Figure 26C).[501–503] In contrast, the prolamins of other cereals, such as wheat, barley, and oats accumulate (along with globulins) in vacuole-derived protein bodies, and in most cases transit through the Golgi complex to arrive there. Rice seeds accumulate two types of storage proteins — the globulin-like glutelins and the prolamins; the former are transported to vacuoles via the Golgi complex, whereas the latter aggregate within the ER lumen.[502,504] These different pathways exhibited by the two storage protein types appear to be facilitated by a differential targeting of their respective mRNAs onto distinct rough ER membranes.[505] During the period of storage protein deposition and protein body biogenesis, the endosperm cells exhibit two morphologically distinct populations of ER membranes: the cisternal ER and the ER that delimits protein bodies (protein body ER). Analyses of the spatial distribution and densities of the specific mRNAs in subcellular fractions enriched in these two ER types reveal a nonrandom distribution. Notably, prolamin transcripts are enriched on the rough ER that delimits the prolamin protein bodies, whereas glutelin mRNAs predominate on the cisternal ER.[505] The molecular basis for segregation of prolamin transcripts to the surface of the protein body ER is presently unknown, although specialized signal peptides and/or components involved in their recognition (i.e., SRPs or ER membrane proteins) have been suggested to play a role.[489,505]

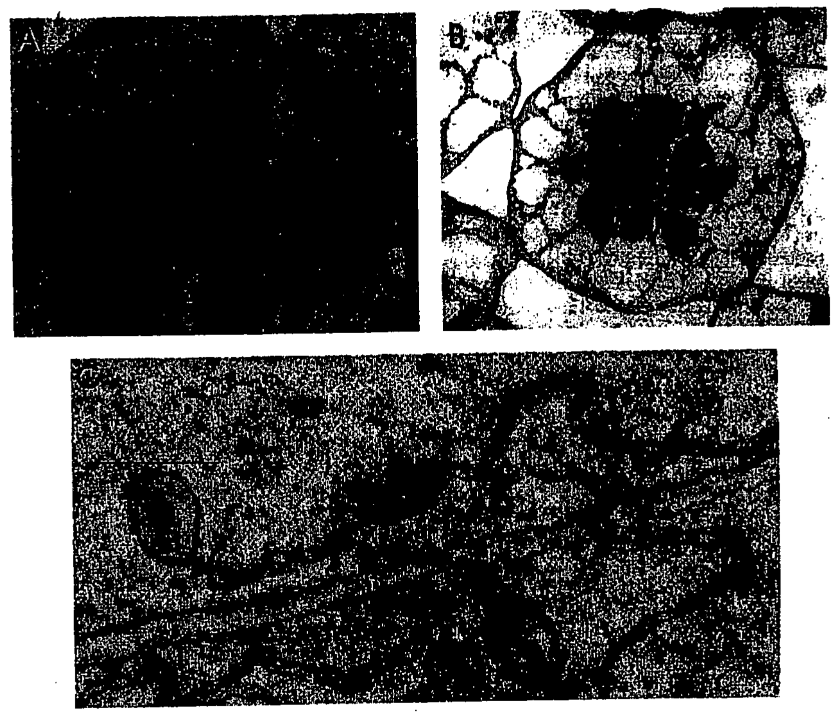


FIGURE 26. Vacuoles and protein bodies in vegetative and seed tissues. (A) Vacuoles within the leaves of *Sophora japonica* trees. The vacuoles are partially or completely filled with protein, including the well-characterized leaf lectins. N, nucleus; C, chloroplast; V, vacuole. For details, see Herman et al.[493] (From Herman, E. M., Hankins, C. N., and Shannon, L. M., *Plant Physiol.*, 86, 1027, 1988. With permission of the American Society of Plant Physiologists. Courtesy of E. M. Herman.) (B) Electron micrograph of a cell from a developing pea cotyledon. Numerous protein storage vacuoles are visible. Storage protein is being accumulated as electron-dense deposits at the periphery of the PSVs. At the end of seed development, PSVs will be completely filled with protein. The nucleus is at the center, surrounded by large, electron-dense, starch grains. (From Vitale, C. and Chrispeels, M. J., *Bioessays*, 14, 151, 1992. With permission from ICSU. Courtesy of S. Craig, CSIRO Plant Industry Division.) (C) Protein bodies originating from the rough ER. Electron micrograph of a developing cell of the starchy endosperm of maize to illustrate the formation of protein bodies (PB) forming from rough endoplasmic reticulum (RER) and the proximity of membrane-bound polyribosomes (MBP). Note the continuity between the protein body membranes and extended RER cisternae (arrow). CW, cell wall. See Larkins and Hurkman.[501] (From Larkins, B. A. and Hurkman, W. J., *Plant Physiol.*, 62, 256, 1978. With permission from the American Society of Plant Physiologists. Courtesy of B. Larkins.)

Another intriguing possibility is that elements of the cytoskeleton are involved. There is a close association of actin filaments with the prolamin-type protein bodies of maize,[506] and a role for the cytoskeletal framework in protein synthesis is firmly established.[507] F-actin may facilitate prolamin-protein body biogenesis by facilitating interaction of prolamin mRNAs with the protein body-

ER. Nonrandom deposition of prolamins may also occur within the endosperm of developing maize, where specific prolamin types are inserted sequentially into protein bodies; correct assembly of different types of prolamins may be required for the formation of normal protein bodies (see Reference 489 and references therein). For example, when individual maize prolamins are expressed in *Xenopus* oocytes, the protein bodies that form have a lower density than that of protein bodies containing all prolamin types (see subsequent discussion, Section VIA, and Reference 602). Although maize prolamin is normally accumulated in ER-derived protein bodies, when expressed in seeds of a heterologous plant host (tobacco) that exhibit protein body biogenesis from subdividing vacuoles, the protein is transported successfully to the vacuole.[507a] Thus, in some cases, maize prolamin is able to escape the normal retention process (but see also Reference 507b).

What, then, is the mechanism of retention of storage proteins within the ER? Is it due to their intrinsic structure or physical properties, or do specific ER-factors also play a role? The conventional mechanism for retention of soluble proteins within the ER is the carboxy-terminal tetrapeptide sequence KDEL or HDEL. Yet, it is unlikely that the mechanism involves this specific retention signal because it is not present on any of the storage proteins that are retained in the ER. Evidence suggesting that the retention mechanism is partly a function of the mature protein comes from expression studies utilizing heterologous hosts (e.g., yeast or *Xenopus* oocytes), in which cereal prolamins are efficiently retained. For example, injection of zein mRNAs into *Xenopus* oocytes results in the synthesis, processing and accumulation of the storage proteins within the ER and assembly into membrane-enclosed structures with the physical characteristics of protein bodies from cereals.[508,509] Failure of prolamins to be secreted from animal cells is not due to improper recognition/processing of the signal peptide or inefficient sequestration into the membrane of the rough ER.[510]

In rice seeds, specific factors in the ER appear to play a pivotal role in protein retention. Here, the ER molecular chaperone BiP retains prolamins within the ER lumen by promoting the folding and assembly of prolamins into protein bodies.[504] Within developing endosperm cells, BiP is localized on the surface of aggregated prolamin protein bodies; dissociation of BiP from these bodies is an ATP-dependent process. It is proposed that biogenesis of protein bodies is a sequential process mediated by BiP. According to this model, BiP binds to the nascent prolamin as it emerges through the ER membrane, an interaction that serves to maintain the polypeptide in a competent state for subsequent assembly into protein bodies. Following the completion of protein synthesis, the prolamin-BiP complexes are released into the ER lumen, and BiP utilizes ATP to assemble the prolamins onto the protein body surface; dissociated BiP is then recycled. A critical concentration of BiP-prolamin complexes within the ER lumen may trigger the initiation of the BiP-mediated aggregation process. Interestingly, BiP has a higher affinity for rice prolamins than it does for rice glutelins, possibly due to the greater proportion of aliphatic amino acids in the former.[511] Although BiP likely associates transiently with glutelins within the ER lumen, it is not localized to the surface of the cisternal ER where glutelins are synthesized.[504,505] The involvement of BiP in prolamin protein body formation is also supported by its elevated concentration in the protein bodies of maize mutants defective in zein accumulation.[137] It is not yet clear whether this mechanism involving a specific ER-associated component (BiP) operates for the retention of other cereal storage proteins that accumulate in ER-derived protein bodies. Retention mechanisms may also involve the physical characteristics of the protein, preventing subsequent transport following translocation into the ER.

The structures of cereal prolamins have been characterized in detail (reviewed in Reference 489; see references therein).[512,513] Most of the prolamins contain several small amino acid repeats that are rich in glutamine and proline. In the sulfur-poor prolamins of wheat, barley and rye, and the α-zeins of maize, these repeats comprise almost the entire length of the polypeptide. In contrast, the sulfur-rich (S-rich) prolamins and the high-molecular-weight prolamins of wheat, barley, and rye contain the repetitive region, but, in addition, contain a 'unique-sequence' globular

region at their C-terminus, which is enriched in α-helices and contains several intramolecular disulfide bonds. Expression of the wild-type S-rich γ-gliadin from wheat and two deletion mutant forms of the protein (missing either the N-terminal or C-terminal regions) in *Xenopus* oocytes has provided some information regarding the possible role of the different regions in protein sorting.[514] When intact γ-gliadin protein is synthesized in *Xenopus* oocytes, a certain proportion of the protein is retained within the ER, while some of the protein is secreted. A deletion mutant composed of only the N-terminal repetitive region of γ-gliadin is retained and packaged into ER-derived protein bodies. Conversely, presence of the C-terminal region leads to efficient secretion from *Xenopus* cells. Thus, sorting of wheat γ-gliadin is proposed to be determined by a balance between two opposing signals: (1) the effectiveness of the N-terminal repetitive region that is responsible for ER retention and assembly into protein bodies within the ER, and (2) the counteracting effect of the C-terminal 'unique-sequence' region that averts packaging within the ER and renders the polypeptide competent for export to the Golgi complex.[514] Moreover, these studies suggested that the retention signal is probably related to the secondary structure of the repetitive region, rather than to its primary structure,[489] although exposed glutamine residues as well as other amino acids present on S-rich prolamins may contribute to the retention process through hydrophobic and hydrogen-bond interactions.

Different retention mechanisms may operate for the maize prolamins, specifically β-, γ-, and δ-zeins, that lack the N-terminal repetitive domain characteristic of maize α-zeins. Localization of γ-zein and truncated forms of γ-zein in *Xenopus* oocytes was determined to examine the role of cysteine-rich domains in ER retention.[515] One truncated zein was deleted in a 94-amino acid C-terminal Cys-rich domain, and the other lacked a 21 amino acid Pro-X linker region that links the Cys-rich domain to a Pro-rich domain. Both truncated forms (as well as intact γ-zein) were efficiently accumulated in *Xenopus* oocytes, indicating that structural features derived from disulfide bonds are not necessary for the retention mechanism of these proteins.[515]

Thus, retention in ER-derived protein bodies in cereals may involve multiple mechanisms, and these mechanisms may interact (e.g., BiP-mediated assembly/aggregation and physical characteristics of the protein, preventing transport out of the ER).

3. Evidence for an Alternative Routing Mechanism to Vacuole-Derived Protein Bodies in Cereal Seeds that Bypasses the Golgi Complex

The transport of prolamins to vacuoles has been studied most extensively in the endosperm of developing wheat grain. In this system, Golgi-derived vesicles contain prolamins, indicative of a route to the vacuole that involves transport through the Golgi complex. However, there is now evidence for an alternative route of prolamin transport to the vacuole that bypasses the Golgi complex. Thus, a considerable proportion of the wheat prolamins assemble into protein bodies within the rough ER. However, there is a subsequent transport of these intact protein bodies to the vacuole where they become internalized by a process analogous to autophagy (Figure 27) (reviewed in References 489, 516, and 517). This process appears to be initiated as protein bodies disconnect from the cisternal ER, an event sometimes accompanied by rupture of the ER membrane. Small electron-lucent vesicles (of unknown origin) then attach and encircle the surface of the protein body, and fuse forming a vacuole containing a protein-body inclusion. In many cases protein bodies inside vacuoles are surrounded by one or two membranes in addition to the vacuolar membranes, suggesting that internalization may occur by an autophagy-like process.[517] Interestingly, a wheat BiP homolog is present within the protein bodies in the cytoplasm as well as inside vacuoles, indicating a Golgi-independent route. It has yet to be determined whether the two routes of transfer of prolamins also operate in other cereals, such as barley and oats. Prolamins of barley may exhibit similar pathways; in the developing endosperm, the constituent protein bodies are surrounded by membranes inside vacuoles (see Reference 489 and references therein). Likewise,

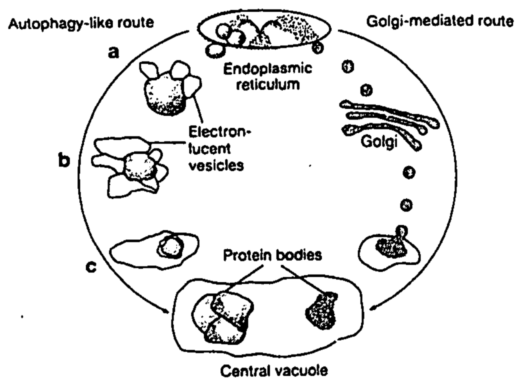


FIGURE 27. Schematic representation of the two different routes by which wheat prolamins transit to vacuoles. Right: The paradigm route, including vesicular transport from the ER via the Golgi complex to vacuoles. Left: Protein bodies formed within the ER become surrounded by electron-lucent vesicles and are internalized into vacuoles by a process analogous to autophagy. (From Galili, G., Altschuler, Y., and Levanony, A., *Trends Cell Biol.*, 3, 437, 1993. With permission from Elsevier Publishing, Cambridge.)

despite the general accumulation of oat prolamins in conventional vacuole-derived protein bodies, some oat prolamins are also detected in protein bodies surrounded by rough ER.

4. Expression of Storage Proteins in Heterologous Hosts: Universality of Targeting Machinery

Because there is such a striking conservation of both general and specific features of the secretory pathway in eukaryotic cells, a pertinent question that has been addressed is whether plant vacuolar targeting signals are correctly processed in other eukaryotes. Plant vacuolar proteins expressed in animal cells (e.g., monkey COS cells, *Xenopus* oocytes) and insect cells, generally undergo normal cotranslational and posttranslational processing that includes signal peptide cleavage and glycosylation; however, they are not retained in any subcellular compartment, but rather are secreted by the bulk-flow pathway.[510,518–520] This further indicates that signals for targeting proteins beyond the Golgi apparatus in plant cells are different from those in animal cells. Not surprisingly, secretion is the outcome only for plant storage proteins that normally transit to their target organelles (vacuoles) via the Golgi complex; as mentioned previously, proteins exhibiting the other major mechanism of transport (i.e., involving deposition into the ER, and subsequently aggregation and formation of the protein body) are retained when expressed in animal cells. Expression of the wheat storage proteins, α-gliadin and γ-gliadin in *Xenopus* oocytes leads to their partial secretion via the Golgi complex; however, secretion occurs at considerably different rates (i.e., at a much faster rate for γ-gliadin), indicative of the different pathways transited by these related wheat storage proteins.[510,514] Some of the features of the plant vacuolar transport and targeting machinery are similar to those of yeast. In those cases studied, when storage protein genes (e.g., PHA) are expressed in yeast cells, the majority of the protein is transported to the vacuole.[521] Some features of vacuolar targeting, then, are conserved in these two highly divergent species. However, although yeast has been used as a model system to attempt to identify plant vacuolar targeting signals, these attempts have largely met with failure. Furthermore, protein sequences or domains on plant proteins that lead to vacuolar targeting in yeast do not function in the same manner in plants and vice versa. Thus, although some components of the targeting pathway are conserved between animals, plants, and yeast (e.g., GTP-binding proteins), other components (e.g., targeting receptors) may be unique.[488]

5. Vacuolar Targeting in Yeast: Genetic and Biochemical Studies

The yeast vacuole, like the plant cell vacuole, is an acidic compartment that functions as a lytic site; it also serves as a storage depot for amino acids, phosphate, and inorganic ions. Numerous soluble hydrolytic enzymes reside in this organelle, including carboxypeptidase Y (CPY), proteinase A (PrA) and proteinase B (PrB). Its limiting membrane also contains a number of proteins and protein complexes, including α-mannosidase, alkaline phosphatase (AlP), dipeptidyl amino-peptidase B (DPAP B), a proton-translocating ATPase, and several permeases.

The soluble vacuolar hydrolases of yeast have been well characterized in terms of their biosynthesis. These proteins are synthesized as inactive precursors (preproproteins) with N-terminal signal peptides that are proteolytically removed during translocation into the ER. Core glycosylation occurs in the ER; additional carbohydrate modifications take place during their transit through the Golgi complex. Sorting of vacuolar enzymes occurs in a late Golgi compartment. Coincident with or just prior to their arrival in the vacuole, the modified precursor proteins (e.g., proCPY, proPrA, and proPrB) are activated by the proteolytic removal of a propeptide segment, a maturation process dependent upon the hydrolase PrA.[522] Considerable progress has been made in characterizing the vacuolar sorting signals of these soluble hydrolases (reviewed in References 364, 523, and 524). In the absence of positive sorting signals for directed targeting away from bulk flow, proteins traversing the yeast secretory pathway are delivered to the cell surface. This fact was exploited to design a screen for mutations in the CPY structural gene (*pcr1*) that result in secretion of CPY.[525,526] Another approach was to construct a nested set of carboxy-terminal deletions in CPY fused to invertase and to follow the fate of the chimeric proteins.[527] This approach demonstrated that the amino-terminal 50 amino acids of the CPY preprodomain, which includes a 20-residue signal peptide, are sufficient to sort invertase protein to the vacuole, while the first 30 amino acids direct high levels of secretion. Deletion analysis of the CPY gene (*pcr1*) in the propeptide-encoding region implicated a domain in the vicinity of amino acid 28 that is essential for efficient sorting of pro-CPY to the vacuole.[526] Deletions in this domain cause the mutant protein to bypass the Golgi sorting reaction and result in mislocalization of most of the proCPY to the cell surface. The analysis of point mutations in the *pcr1* gene, encoding amino acid residues Lys 18 (in the signal sequence) to Leu 34, identified only four contiguous residues important for vacuolar sorting, $LQRP_{26}$ (Leu-Gln-Arg-Pro).[526,528] Thus, this tetrapeptide sequence constitutes the core of an amino-terminal topogenic element, immediately following the signal peptide, that is necessary for targeting proCPY to the yeast vacuole.[523] Interestingly, the context in which the CPY tetrapeptide sequence is presented affects the efficiency of targeting, inferring the involvement of secondary structural elements in the sorting mechanism of this protein.[528]

Subsequent analysis of the targeting signals of other vacuolar soluble hydrolases (e.g., PrA and PrB) revealed that the sorting mechanisms in yeast are much more complex than the initial simplistic model that emerged from the CPY studies. The sorting domain of these proteins is unlikely to be a short linear amino acid sequence. For PrA it may be encoded by two signals, one in the mature protein and the other in the propeptide domain (e.g., between amino acids 61 and 76).[529] There is no sequence similarity between the propeptides of CPY and PrA. Thus, currently no consensus sequence or common structural determinant has been demonstrated for yeast vacuolar

proteins, suggesting that a diverse array of factors may be operative in the sorting process. Although not clearly defined, the sorting domain of PrB likely resides in the mature portion of the protein because the precursor undergoes an unusual early processing event in which the 280 amino acid propeptide is removed in the ER.[530]

Although some commonalities exist, there are independent mechanisms for the delivery of soluble and membrane proteins to the yeast cell vacuole. This appears to be the case for sorting of the vacuolar (type II) integral membrane protein AIP.[531] Transport of this protein is less sensitive to some of the defects of the vacuolar sorting apparatus in mutants that exhibit dramatic missorting and secretion of the soluble hydrolases (e.g., CPY and PrA; see subsequent discussion). DPAP B also resides in the yeast vacuole as a type II integral membrane protein having short (amino-terminal) cytoplasmic and transmembrane domains (of 45 amino acids in total) and a large (800 residues) luminal domain.[532] No single domain of DPAP B is required for delivery to the vacuolar membrane. Removal or replacement of either the cytoplasmic transmembrane or luminal domain does not affect the protein's transport to the vacuole.[533] DPAP A normally resides in the yeast Golgi complex; both overproduction of this protein and a mutation within its cytoplasmic domain result in mislocalization to the vacuole. These results have invoked a model in which membrane proteins are delivered to the vacuole along a default pathway.[533] Interestingly, α-mannosidase, a marker enzyme of the vacuolar membrane in yeast, is delivered to the vacuole by a novel route separate from the secretory pathway.[534]

It is likely that yeast vacuolar protein targeting signals are recognized by a sorting receptor(s), and the resultant receptor-ligand complexes are subsequently sorted into vesicles and transported to the vacuole by additional constituents of the vacuolar sorting machinery. Toward identifying the components of this sorting process, the application of several genetic selections has resulted in the isolation of a number of mutants that exhibit defects in vacuolar protein localization (primarily of the soluble hydrolases) and/or processing (reviewed in Reference 523). Instead of delivering vacuolar hydrolases to the vacuole, these *vps* (vacuolar protein sorting defective) mutants missort the enzyme precursors to the yeast cell surface.[523,535–537] Posttranslational glycosylation and secretion of proteins is normal in most mutants; thus, their defects are specific for targeting to the vacuole.[535,536] Genetic comparisons among the *vps* mutants demonstrate that they collectively define more than 50 unique complementation groups, indicating that delivery of proteins to the yeast cell vacuole represents a highly complex process requiring the coordinated participation of a relatively large number of gene products. Some of these components are likely to include sorting receptors and proteins involved in the formation, interorganelle transport, targeting, and fusion of vesicles. Biochemical characterization of the *vps* mutants should not only identify cellular components directly involved in the specific segregation, packaging, and delivery of vacuolar proteins, but also those gene products involved in the regulation or control of these processes.

Progress toward this goal has been achieved recently with the characterization of protein products of some *vps* mutants (Table 9). The VPS1 protein is homologous to a group of proteins with a GTP-binding motif, some of which also bind microtubules.[538] Another VPS protein, VPS15, is a serine/threonine kinase that is essential for the delivery of soluble hydrolases to the vacuole.[539,540] The VPS15 protein is preferentially associated with the cytoplasmic face of a late Golgi or vesicle compartment and may regulate specific protein phosphorylation reactions required for efficient delivery and sorting of proteins to the yeast cell vacuole. For example, mutational alteration of the VPS15 protein kinase domain results in the biological inactivation of this protein and the secretion of multiple vacuolar hydrolases. Protein phosphorylation has been implicated as a key regulator of protein sorting, specifically acting within pathways at branch positions, where proteins must choose between two or more different transport fates.[539,541,542] Thus, the proposed role of protein phosphorylation is as a molecular "switch" within intracellular transport pathways such that proteins are actively diverted from a default pathway (i.e., secretion) to an alternative pathway (e.g., transport to the vacuole). In carrying out its role,

TABLE 9
Genes that May Function in the Transport and Sorting of Proteins to the Yeast Vacuole[a]

Gene	Phenotype	M_r (kDa)	Sequence features	Protein features and localization[c]	Protein function *in vitro*
vps1	Golgi membrane, multivesicular bodies; secretes vacuolar proteins	80	Identity with *S.c.* SPO15, similar to *D.m.* shibire[b]	Punctate localization	GTP-binding
vps3	Secretes soluble vacuolar proteins	140	—	Partially soluble protein	—
vps5	Fragmented vacuole; secretes soluble vacuolar proteins	90	—	Partially soluble protein, phosphorylated	—
vps15	Secretes soluble vacuolar proteins; no effect on membrane proteins	166	Catalytic domain of Ser/Thr kinases, type II phosphatase	Peripheral membrane protein, N-terminal myristolation, autophosphorylated; fractionates with late Golgi marker	—
vps17	Fragmented vacuole; secretes soluble vacuolar proteins	70	—	Peripheral membrane protein, phosphorylated	—
vps33	Vacuole absent; secretes vacuolar proteins	75	ATP-binding motifs	Soluble protein	—
vps34	Secretes vacuolar proteins	95	—	Partially soluble protein, punctate distribution, phosphoprotein, coimmunoprecipitates with VPS15	Phosphorylation

[a] Many VPS proteins have not been characterized yet.
[b] *S.c.* = *Saccharomyces cerevisiae*; *D.m.* = *Drosophila melanogaster*.
[c] Proteins designated peripheral membrane may also exist in soluble, cytoplasmic form; proteins designated partially soluble sediment with a Triton X-100-insoluble fraction.

Based on Pryer, N. K., Wuestehube, L. J., and Schekman, R., *Annu. Rev. Biochem.*, 61, 471, 1992. With permission from The Annual Review of Biochemistry, by Annual Reviews, Inc.). See references therein.

the VPS15 protein appears to functionally interact with another protein that is encoded by the gene *vps34*.[539,543] The VPS34 protein is a phosphatidylinositol 3-kinase involved in the formation of 3-phosphorylated phosphoinositol-phosphates and known to interact with activated cell surface receptor tyrosine kinases.[544] VPS34 is thought to be present in an unknown intermediate organelle where sorting to the vacuole may occur. Point mutations altering highly conserved residues within the VPS34 kinase domain result in inactivation of VPS34p/PI3-kinase and missorting and secretion of vacuolar proteins, confirming the role of this protein in regulating intracellular protein trafficking decisions. The presence of phosphoinositides in the lipid bilayers of organelles and vesicles may allow fusion of transport vesicles with their target membranes or enable binding of cytosolic and/or cytoskeletal proteins to transport vesicles (reviewed in Reference 396). Recently, a plant homolog of the yeast *vps34* gene has been isolated.[545]

The mechanism or level of control involving phosphorylation may be superimposed upon that carried out by GTP-binding proteins (viz., mediating the unidirectionality of secretory protein traffic and designating a transport vesicle's final destination; see earlier discussion).[539] Three

genes, *ypt51*, *52*, and *53*, have been isolated from *S. cerevisiae* that encode small GTPases with extensive homology to mammalian Rab5 and the *S. pombe* YPT5 protein.[546] The defects associated with single, double, and triple null mutants of *ypt51*, *ypt52*, and *ypt53* suggest an important function of the encoded proteins in the delivery to the vacuole of two endocytic markers, α-factor and lucifer yellow CH. Moreover, a number of vacuole-related defects associated with the *ypt* mutants such as protein sorting, acidification, and morphological changes, are indicative of a key role for these YPT proteins in vacuole biogenesis.[546] In developing pumpkin cotyledons, small GTP-binding proteins are associated with dense vesicles targeted to vacuoles. Studies are underway to determine whether these proteins play a pivotal role in vesicle targeting and/or fusion to the plant vacuolar membrane.[547]

Vps16 gene function leads to severe defects in the sorting of soluble and membrane-associated vacuolar proteins and greatly perturbs vacuole morphology.[548] Fractionation studies indicate an association of the VPS protein with a large proteinaceous complex that is required for vacuole biogenesis and/or vacuole stability. It may also associate with a limited number of sites on cytoskeletal elements.

The generation and maintenance of a low luminal pH within organelles of the secretory pathway is critical for vacuolar sorting (reviewed in References 523 and 524). Specifically, acidification is presumed to play a role in promoting the dissociation of vacuolar proteins from their sorting receptors, allowing the unbound receptors to return to the Golgi complex. Agents that abolish acidification of the vacuolar lumen (e.g., ammonium chloride, a classic lysosomotropic agent, or balifinomycin A1, a specific and potent inhibitor of vacuolar H^+-ATPases) also promote mislocalization of newly synthesized vacuolar proteins to the cell surface.[549,550] Presumably, these agents result in the recycling of bound receptors, causing a depletion of the pool of receptors available for sorting; secretion of newly synthesized vacuolar proteins by bulk-flow transport is the outcome. The yeast vacuolar H^+-ATPase is a multimeric enzyme containing at least 8 subunits of 100, 69, 60, 42, 36, 32, 27, and 17 kDa (see References 551 through 554 and references therein). Most of the structural genes encoding these subunits have been cloned. The 69- and 60-kDa proteins bind ATP and are thought to form the catalytic site, whereas six copies of the 17-kDa protein are thought to form the proton channel (reviewed in Reference 555); other proteins participate in the assembly and localization of the vacuolar H^+-ATPase complex.[523,551,553,554] For example, the protein Vma21 is an ER membrane protein required for assembly of the vacuolar H^+-ATPase complex.[551] The 42- and 27-kDa subunits are essential for assembly of the peripheral membrane portion of the H^+-ATPase onto the vacuolar membrane.

The analyses of plant cells defective in H^+-ATPase activity (pH mutants) will allow a direct test of the role of acidification in the sorting of plant vacuolar proteins.[556]

Future studies in yeast will include a detailed characterization of the *vps* genes and their encoded products. The use of *in vitro* reconstitution assays for vacular protein sorting[557] should aid in revealing the precise roles of these proteins in the sorting process (e.g., in receptor-mediated recognition/sorting, vesicularization and vesicle targeting, and fusion events). A clear understanding of these roles should in turn provide insight into how the various sorting components interact as well as the more general roles of protein phosphorylation and GTPases in intracellular protein sorting processes. Although yeast has not been a suitable model system for identification of plant vacuolar targeting signals, it may be an appropriate model system to identify novel mechanisms of protein retention within the plant ER.[489] A yeast system would be particularly valuable to study targeting of plant proteins, given its extensive set of secretion and BiP mutants. Moreover, transformation techniques for cereals are currently limited, and long periods of time are required to obtain transgenic cereal plants.[489]

6. Identifying Plant Vacuolar Targeting Signals

a. Use of Transgenic Yeast as the Heterologous Host System

Because the plant vacuolar protein PHA is correctly processed and sorted to the yeast cell vacuole,[521] transgenic yeast has been used as a model system to attempt to define the vacuolar targeting signal on this protein.[558] A PHA fusion protein containing 43 amino-terminal residues of the mature protein, together with the signal peptide, is sufficient to redirect the secreted form of yeast invertase to the yeast vacuolar compartment. Deletion analysis further localized the vacuolar sorting domain to an amino-terminal portion of mature PHA, between amino acids 14 and 23.[558] Interestingly, this domain contains a yeast-like vacuolar targeting (tetrapeptide) sequence, $LQRD_{21}$, reminiscent of the $LQRP_{26}$ sequence on CPY, and conserved to some degree in other lectin proteins of legumes.[17] Site-directed mutagenesis to effect amino acid changes within this tetrapeptide sequence in PHA-invertase fusion proteins has demonstrated the importance of this short domain in the vacuolar sorting reaction in yeast. In particular, exchanging the aspartate at position 21, which introduces a site for addition of a glycan into the sequence, results in considerable secretion of invertase activity (e.g., 64%). However, the effect of similar mutations within this short domain in full-length PHA does not significantly affect targeting of the protein to the vacuole, indicating that there is additional vacuolar targeting information in PHA.[17,558] Thus, a second independent signal present towards the middle of the polypeptide (i.e., carboxy-terminal to the first domain identified) may also be essential for correct sorting in yeast cells, notably similar to the requirements for PrA and PrB (see earlier discussion).[17]

Perhaps more important, however, is the observation that the PHA-invertase fusion proteins that direct transport to the yeast cell vacuole are not targeted successfully to vacuoles in *A. thaliana*. Therefore, the sorting determinant that contains sufficient information for vacuolar targeting in yeast lacks the necessary information for efficient targeting in plants, suggesting that vacuolar sorting signals in these two organisms are dissimilar, perhaps in the extent to which other determinants in the mature protein are required for their receptor-mediated recognition.[17] Further studies on various PHA-invertase chimeric constructs expressed transiently in plant protoplasts reveal that vacuolar sorting information is contained within an internal domain of PHA that is predicted to be exposed at the surface of the folded molecule.[559] The internal domain consists of 30 amino acids (amino acids 84 to 113 in mature PHA) and is capable of directing 50% of the reporter protein to the plant vacuole. It does not appear to share any sequence homology with the vacuolar targeting signals identified on other plant proteins (see discussion below).

Fusion proteins of pea legumin and yeast invertase indicate that targeting information (for the yeast vacuole) is contained in both the amino-terminal and carboxy-terminal portions of the legumin protein.[560] The vacuolar sorting signals of some yeast and possibly certain plant proteins (e.g., legumin of field bean)[560] may be composed of regions on the surface of the protein that are commonly termed 'signal patches' (Figure 2). Unlike signal peptides, signal patches will, in general, be formed from noncontiguous regions of the polypeptide chain that are brought together during protein folding; thus, they are conformation dependent.[16] A dependence on signal patches for correct sorting would obviously contribute to the difficulty of defining these domains experimentally, particularly via engineered (chimeric or deletion) proteins that contain (or lack) only continguous amino acid sequences (Figure 3). Although we presume it to be the case, we do not yet know with certainty whether the three-dimensional conformation of proteins is precisely the same in the heterologous yeast vs. plant-host cells; there may be subtle differences, such that different protein determinants or features are more (or less) accessible for receptor-mediated recognition in these two systems. However, the sorting determinants per se that are critical for plant vs. yeast vacuolar targeting appear to be different.[488,561,562] For example, as will be discussed in the next section, a carboxy-terminal propeptide (CTPP) on barley lectin is a vacuolar targeting signal in plants, being both necessary and sufficient for vacuolar localization in transgenic tobacco. Barley lectin domains redirect the normally secreted reporter protein invertase, for sequestration in yeast, but they do so in a manner independent of the carboxy-terminal propeptide.[561] Invertase with only the plant protein carboxy-terminal propeptide is secreted in yeast. Conversely, two fusion proteins consisting of invertase linked to barley lectin containing its carboxy-terminal propeptide or to

barley lectin lacking this amino acid-stretch are both retained in yeast cells.[561] Similar results have been obtained with another plant protein, sporamin, in which the plant vacuolar targeting signal is a N-terminal propeptide.[562] Thus, none of the three types of plant vacuolar-targeting signals identified so far (i.e., an internal domain, a N-terminal propeptide, or a C-terminal propeptide) are recognized in yeast. Moreover, in all cases, another cryptic signal was found, confirming the lack of conservation of certain components of the vacuolar targeting machinery between plants and yeast.

b. Use of Transgenic Plants as the Heterologous Host System

Several studies have demonstrated that plant vacuolar proteins are correctly targeted in other plant hosts (reviewed in Reference 17). In general, the vacuolar targeting signals on these proteins are recognized with a high degree of fidelity, regardless of the species, cell type, or organ in which expression is directed. However, transport efficiency may be variable and has not been systematically determined; likewise, there may be differences between vegetative and seed tissues (see later discussion). Progress has been made toward identifying the molecular mechanisms regulating the vacuolar sorting of Gramineae lectins.[563-566] Barley lectin is synthesized as a preproprotein with a glycosylated CTPP that is removed just prior to, or concomitant with, deposition of the protein in vacuoles.[567] The intact protein is correctly assembled, processed, and targeted to the vacuole in heterologous tobacco host cells.[82] Expression of a mutant form of the lectin gene, in which the region encoding the short pro-domain is deleted, results in missorting and secretion of the protein via the bulk-flow transport pathway.[563] The CTPP on the lectin is also sufficient for vacuolar targeting; when the corresponding DNA sequence is fused to a gene encoding a secreted reporter protein (cucumber chitinase), the resultant fusion protein is redirected to the vacuole in transgenic tobacco with 70 to 75% efficiency.[464,490] Other Gramineae lectins such as wheat germ agglutinin and rice lectin contain similar proteolytically processed C-terminal domains,[568] and these probably serve the same function. Sporamin (a vacuolar storage protein in tuberous roots of sweet potato) undergoes a pattern of proteolytic processing that is similar to that of the cereal lectins. This protein is synthesized as a preproprotein with a short amino-terminal propeptide (NTPP) domain that undergoes cleavage in a posttranslational manner, probably in the vacuole.[569] Studies of the fate of sporamin gene constructs in transgenic tobacco show a dependence on the amino-terminal prodomain for correct vacuolar targeting, similar to that demonstrated for the carboxy-terminal prodomain of barley lectin.[570] Thus, in these two vacuolar proteins (which have a similar pattern of precursor synthesis and posttranslational proteolytic processing) the prodomains appear to contain information that is essential and sufficient for vacuolar sorting.

Subsequent studies analyzing the functional elements of the barley lectin C-terminal prodomain suggest that no specific amino acids are involved; many different alterations and deletions were tolerated by the targeting machinery, although short hydrophobic amino acid stretches seemed important.[566] As few as three amino acids of the CTPP are sufficient for the correct targeting of barley lectin to the plant vacuole. Neither charge nor glycosylation of the prodomain are necessary for targeting. The predicted secondary structure (i.e., an amphipathic α-helix)[571] is also not required for recognition by the vacuolar sorting apparatus; however, correct three-dimensional presentation of the domain appears to be essential.[564,565] Addition of two glycine residues after the terminal glutamic acid (Table 10) or the addition of an N-linked glycosylation site, four amino acids from the carboxy-terminus, causes secretion of barley lectin from tobacco cells.[566] Thus, the carboxy-terminus is clearly the site of recognition by the sorting machinery and the addition of glycine residues or a bulky glycan at this site significantly alters delivery of the protein to the vacuole.[488] Comparison of the propeptide sequence of sporamin to other propeptide sequences (e.g., on barley aleurain and potato cathepsin D inhibitor) reveals a common short region of hydrophobic amino acids (NPIRLP) with a hydrophilic residue, Arg, in the center (Table 10); Asn (N) and Ile (I) are conserved, and Pro (P), Arg (R), and Leu (L) can be substituted.[492,572] Progress is being made toward determining the critical residues in this region that are necessary and sufficient for vacuolar sorting.[572,573] Conserved amino acids in the N-terminal part of the propeptide of sporamin ($N^{26}PIRL^{30}$) (Table 10) may constitute the core of the vacuolar-sorting determinant; both the side-chain structures of the conserved residues and the higher-order structure of the polypeptide backbone in this region are likely recognized by the vacuolar sorting machinery.[573]

Propeptides of vacuolar proteins do not necessarily contain targeting information. A short C-terminal segment of a Brazil nut 2S albumin is sufficient for targeting to the vacuole;[574] conversely, none of the three propeptide segments of the *Arabidopsis* 2S albumin storage protein (which undergo proteolytic removal) are involved in vacuolar localization.[575]

Recent studies with some of the pathogenesis-related proteins of tobacco (e.g., chitinase, β-glucanase) have also led to the identification of the vacuolar sorting domains on these proteins. Homologous extracellular and vacuolar forms of these hydrolytic enzymes exist; acidic forms are generally extracellular, whereas their basic counterparts (likely encoded by different gene products) are vacuolar.[26,576,577] A basic chitinase of tobacco is synthesized as a higher molecular weight precursor, with three domains that are not present on the acidic form (viz., an insertion in the middle of the polypeptide and short amino-terminal and carboxy-terminal domains).[25] The carboxy-terminal tail of the basic tobacco chitinase appears to have significant vacuolar targeting information;[578] deletion of a carboxy-terminal heptapeptide from the basic enzyme leads to secretion of the mutant chitinase from transgenic tobacco cells. Conversely, a fusion protein (containing the carboxy-terminal domain of basic tobacco chitinase linked to an acidic [extracellular] cucumber chitinase) is retained to a substantial degree (i.e., 50%) in transgenic tobacco cells and is sufficient for vacuolar localization. Likewise, the targeting information on the vacuolar antifungal tobacco enzyme AP24 resides in its short carboxy-terminal propeptide that is removed concomittantly, or subsequent to, transport to the plant vacuole.[579]

A comparison of the C-terminal extensions of the lectins and vacuolar hydrolases (tobacco chitinase and β-glucanase) reveals no amino acid identities (Table 10); however, a common feature is an abundance of hydrophobic amino acids, which may be important for recognition by sorting machinery.[490] Sequence changes in the C-terminal vacuolar targeting peptide of tobacco chitinase lead to a gradual transition from vacuolar retention to secretion, suggesting that (at least for some proteins) the vacuolar sorting system has low specificity for the primary sequence of the targeting domain.[580]

Aleurain, a vacuolar thiol protease in barley aleurone layer cells, is made as a 42-kDa proenzyme (proaleurain). Aleurain is proteolytically processed into its mature form by the removal of a N-terminal propeptide[338] whose amino acid sequence shares identities with the sporamin prodomain (Table 10).[83] Two steps are required to form the mature 32-kDa aleurain; the first yields a 33-kDa intermediate; subsequent trimming results in the gradual loss of 1 kDa. Two homologous thiol proteases, aleurain and EP-B, each of which have N-terminal propeptides of about 110 amino acids, are simultaneously expressed in aleurone layer cells, but have different destinations (vacuolar vs. extracellular, respectively). Toward identifying the vacuolar sorting signal on aleurain, chimeric proteins resulting from switching prosequences from the two proteins have been constructed and their fate (i.e., subcellular localization) examined in heterologous host plant cells.[178,581] Substitution of the propeptide of EP-B with the N-terminal propeptide of aleurain (containing the amino acids SSSSFADSNPIR) results in redirection of about 50% of EP-B to the vacuole.[178] Shorter sequences of this domain (SSSSFADS and SNPIR) are also able to target pro-EP-B to the vacuole, albeit with much lower efficiencies compared with the combined sequence. Efficient vacuolar sorting of aleurain may, therefore, be mediated by the combined action of small contiguous determinants.[178,581]

Thus, significant insight into vacuolar sorting may be derived from comparisons of amino acid sequences of homologous extracellular and vacu-

TABLE 10
Vacuolar Sorting Signals in the N-Terminal and C-Terminal Propeptides of Plant Vacuolar Proteins

Vacuolar protein	Location of propeptide	Targeting signal
Sweet potato sporamin	N-terminal	+ H S R F N P I R L P T - - -
Barley aleurain	N-terminal	+ S S S S P A D S N P I R P V T D R A A S T - - -
Barley lectin	C-terminal	- - d g V F A E A I A A N S T L V A E -
Tobacco chitinase A	C-terminal	- - g n G L L V D T M -

Note: In the amino acid sequence of the propeptide, hydrophobic amino acids are indicated by bold letters, and the positive and negative charges in the polypeptides are indicates by + and –, respectively. ∇ indicates the cleavage site of the CTPP. The N-linked glycan (Y) is attached to an Asn residue in the barley lectin CTPP. The exact N-terminal amino acids of prosporamin and proaleurain in tobacco cells are not known. The NPIR motif in the N-terminal propeptides and the hydrophobic/acidic motif in the CTPPs are indicated by the underlines (solid underlines and broken underlines, respectively).

Based on Nakamura, K., and Matsuoka, K., *Plant Physiol.*, 101, 1, 1993. With permission from the American Society of Plant Physiologists.

olar proteins.[17] Subsequent studies utilizing this information can then be geared to identifying those domains that are both necessary and sufficient for correct targeting of the vacuolar form.

In legume root nodules, proteins of the peribacteroid fluid are targeted through the endomembrane system to the peribacteroid membrane. Although the mechanisms for targeting of peribacteroid membrane proteins are presently unknown, they may be various as indicated by the nature of nodulin-26 vs. nodulin-24.[582,583] A contributing factor to targeting of these nodulin proteins may be that the peribacteroid membrane is a mosaic, having properties common to both the plasma membrane and the vacuole.[417]

Although there are exceptions, it is noteworthy to mention here that identification of vacuolar targeting signals on seed storage proteins may not be successful by analyzing the fate (i.e., subcellular localization) of engineered proteins in plant vegetative tissues. This appears to be the case for vicilin of pea seeds.[584] Subcellular localization studies of vicilin deletion mutant and chimeric proteins in leaf cells indicated possible participation of both N-terminal and C-terminal regions of the mature protein in vacuolar targeting; both a 121-amino acid N-terminal deletion mutant and a 69 amino acid C-terminal deletion mutant failed to accumulate in the leaf vacuole. However, protein degradation in the leaf vacuole and disrupted assembly were likely to be important factors contributing to the low levels of deletion mutant proteins detected in this compartment. Highly specific proteolytic processing was exhibited by the same mutant proteins synthesized in tobacco seed that may reflect significant correct targeting to the storage vacuole/protein body in seed tissues. Interestingly, expression of the gene encoding the α-subunit of soybean β-conglycinin in transgenic petunia (under control of the cauliflower mosaic virus 35S promoter) leads not only to differential levels of protein accumulation but also to differential processing and transport in the leaf vs. the seed. For example, the two major products in seeds are 76 kDa (the intact α′ subunit) and 55 kDa (a breakdown polypeptide of the α′ subunit) that accumulate only in protein storage vacuoles (protein bodies). In contrast, a polypeptide of 80 kDa, the predicted size of the α′ precursor, and a 53 kDa breakdown polypeptide are found in leaves and other non-seed tissues.[585] Moreover, in the non-seed tissues, these products were localized within a microsomal membrane fraction containing ER and Golgi. On a cautionary note, this may simply indicate that the products are unstable in the leaf vacuole, and proteins in the ER and Golgi represent the small amounts of protein in transit to the vacuole. In somatic embryos of alfalfa (derived from vegetative tissues), there is mistargeting of 2S albumin storage proteins (normally vacuolar proteins) that localize to the cytosol.[586] Thus, it appears that problems related to protein stability and possible loss of fidelity of targeting make the vegetative tissues inappropriate model systems to identify the vacuolar targeting signals on some seed storage proteins.

c. Targeting of Membrane Proteins to the Tonoplast

Some progress has been made toward identifying the targeting signals of tonoplast proteins in plants.[180,203] As mentioned earlier, in bean seeds, α-TIP is synthesized on the rough ER, and its transport to the tonoplast is mediated by the secretory system. The C-terminal 48 amino acids of α-TIP, which include the sixth membrane-spanning domain and the cytoplasmic tail, can redirect a soluble reporter protein to the tonoplast in tobacco cells. A mutant form of α-TIP (in which the C-terminal cytoplasmic tail is deleted) is still targeted to the tonoplast. The C-terminal transmembrane domain may contain the information for transport to the tonoplast; however, deletion of the sixth membrane-spanning region results in an unstable protein.[203] Recent work in yeast suggests that targeting of membrane proteins to the vacuole occurs via a default mechanism.[533,587] For example, removal of a positive targeting signal from a Golgi-associated membrane protein (KEX1) results in its mislocalization and transport to the vacuolar membrane.[587] In plants, work with the transport inhibitors monensin and brefeldin suggests that soluble proteins of the vacuole (e.g., PHA) and membrane proteins (e.g., α-TIP) reach their vacuolar destinations by different paths.[180] As mentioned earlier, in some eukaryotic sys-

tems, brefeldin A prevents anterograde vesicle transport between the ER and Golgi complex. Specifically, it prevents the assembly of coatomers due to its inhibition of the GTP-dependent interaction of the ADP-ribosylation factor with the Golgi membrane.[588,589] Monensin, on the other hand, inhibits correct sorting in the TGN by disrupting the proton gradient across the membrane. Both inhibitors prevent vacuolar transport of PHA, but are ineffective in blocking arrival of α-TIP in the tonoplast.[180] Although these results are indicative that transport to the plant tonoplast represents a default pathway, further work is necessary to confirm this model, including a determination of the unique features of the plant Golgi complex as well as the precise effects of monensin and brefeldin on plant secretory pathway functions.

d. Other Components of the Plant Vacuolar Targeting Machinery: A Search for Receptors

To fully understand plant vacuolar targeting mechanisms, it will be necessary to undertake a detailed characterization of the sorting signals from various proteins with different functional and structural characteristics. The above studies make it clear that the sorting signals on plant vacuolar proteins are various and include internal domains, 'signal patches,' and prodomains, the latter removed when vacuolar transport is completed (Figure 28). The identification of specific sequences required for vacuolar targeting has initiated a search for receptors mediating the recognition of those signals.[590,591] A comparison of targeting sequences on different vacuolar proteins

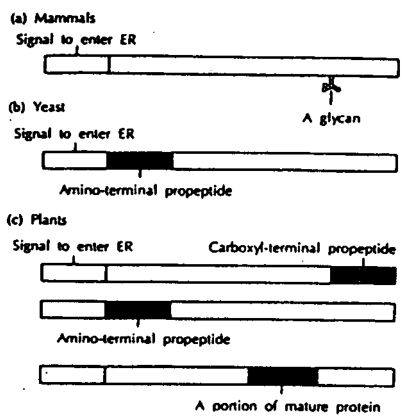


FIGURE 28. A summary of targeting signals of vacuolar/lysosomal proteins in eukaryotes. (c) Shows three different regions in which the vacuolar sorting and targeting information occur in plant proteins: in the cleaved carboxy-terminal or amino-terminal propeptide, or internally within the mature protein. The signal peptide for insertion in the ER is shown for each protein. See Chrispeels and Raikhel.[490] See also Holwerda et al.,[178] Bednarek and Raikhel,[83] and Nakamura and Matsuoka.[492] (From Gal, S. and Raikhel, N. V., *Curr. Opin. Cell Biol.*, 5, 636, 1993. With permission from Current Biology.)

has revealed little amino acid similarity; thus, it is pertinent to ask whether different classes of targeting signals compete for the same receptor[592] (reviewed in Reference 488). Tobacco plants synthesizing barley lectin (containing a CTPP) were crossed with those expressing sporamin (having an NTPP), and the progeny was analyzed. Competition between these two proteins for the same receptor was expected to lead to secretion of one of the proteins; however, both proteins were correctly processed and localized to the same vacuoles.[592] Thus, the carboxy-terminal and NTPPs are equally recognized by the vacuolar protein-sorting machinery indicating the existence of multiple sorting receptors or a broad-specificity sorting mechanism that is not easily saturated.

An integral membrane protein of ~80 kDa (BP-80) has been identified in clathrin-coated vesicles of developing pea cotyledons. The protein binds specifically (at neutral pH) to the N-terminal targeting signal of proaleurain (bound to a column) and dissociates from its ligand when the pH is lowered to 4. The NTPP of sporamin competes (weakly) for binding of the putative receptor protein; a mutant form of this NTPP does not. Likewise, endopeptidase-B (EP-B), a homologous, secreted thiol protease, also does not exhibit an interaction with the 80-kDa protein. The transmembrane orientation of the protein is also indicative of a receptor function; the carboxy-terminus is exposed to the cytoplasm, whereas the binding domain is located in the N-terminal luminal portion of the protein.[591] At present there is no direct functional evidence that BP-80 acts as a vacuolar receptor. However, both pea and *Arabidopsis* have multiple genes for proteins closely related to BP-80 and similar genes are expressed in maize and rice (reviewed in Reference 591a; see references therein). Further characterization of this putative plant vacuolar targeting receptor will be awaited with interest. Another future goal is the isolation of plant mutants having specific defects in vacuolar protein targeting (similar to those in yeast); these will be valuable for identifying other components integral to the transport and sorting machinery.[330] As will be discussed in the next section, this basic research will be of value for more applied studies geared toward the genetic engineering of plants with improved traits.

VI. APPLICATIONS TO GENETIC ENGINEERING: MAXIMIZING LEVELS OF GENE EXPRESSION IN FOREIGN PLANT HOSTS

Subsequent to the advent of genetic engineering of plants, there has been an ongoing thrust toward refinement of the techniques for gene isolation and manipulation, plant host transformation, and efficient regeneration of whole plants from single transformed cells. Current attention is focused on using the technology to define or elucidate regulatory mechanisms of plant gene expression and as a means of introducing agronomically useful traits or characters into crop plants. One major goal is to engineer seed crops for greater nutritional value by introducing storage proteins modified to contain a more optimal balance of the essential amino acids (i.e., those amino acids that cannot be synthesized by animals and hence, must be supplied in their diet). Another related goal (particularly important to the economy of Australia and New Zealand) is to obtain high levels of accumulation of certain sulfur-rich, rumen-resistant storage proteins within the leaves of pasture plants (e.g., lucerne and subclover), with the aim of increasing wool growth in sheep.[156,157] Other desirable traits that researchers are presently focused on include the engineering of insect, disease, and herbicide resistance in crop plants, as well as the production of medicinally important substances in plants.

To be of agronomic value (particularly in relation to enhanced nutritive value), levels of the foreign protein must accumulate to significant levels in the host plant (e.g., to about 1 to 10% of the total cellular protein); it may also be desirable to target protein accumulation into a suitable subcellular compartment. The former objective is particularly challenging because levels of foreign proteins in transgenic plants (particularly within vegetative tissues) are generally low, although there are notable exceptions.[593-596] Thus, concerted efforts are being made to enhance gene transcription and mRNA stability and translatability to achieve high protein levels. Such attempts generally have been successful,[156,157] yet in some instances have led to insufficient increases in foreign protein accumulation to be of economic consequence. More recently, a novel strategy has

been the "mis-" or "retargeting" of proteins into different subcellular compartments to improve the stability of foreign proteins in transgenic plants.

A. Enhancing Protein Stability in Transgenic Host Plants

It has become evident that a lack of stability of the introduced protein in the foreign host environment is a major obstacle that must be overcome to obtain high levels of protein accumulation. Protein instability may be due to structural constraints (e.g., as a consequence of mixed or additional coding sequences); it may also arise as a consequence of the subcellular localization of the protein in the foreign host environment. Thus, success in achieving high protein levels in transgenic host plants will likely depend on a close scrutiny of targeting signals, as well as the structural features of proteins that allow for their stable accumulation in cells. The importance of these factors is underscored by studies attempting to introduce storage proteins that have been structurally modified to enhance their nutritional quality.[311,597] For example, a high-methionine β-phaseolin gene (modified to contain codons for 15 additional amino acids, including six methionine residues), yielded very low levels of accumulation in transgenic tobacco seeds, as a consequence of increased protein degradation. The majority of the modified protein was degraded within the seed vacuole.[311,312] Similarly, a gene with a modified CAT (chloramphenicol acetyl transferase) coding region (in which a 292-base-pair coding sequence encoding 80% essential amino acids was inserted) gave very low levels of protein in transgenic potato tubers.[597]

Knowledge of the rules that relate amino acid sequence to protein structure is not only fundamental to our understanding of such important biological processes as protein folding and the achievement of a functional three-dimensional structure, but also will be essential to any attempts to modify the nutritional quality of proteins. Tolerance to amino acid substitutions in relation to protein structure (and hence stability) and function is just beginning to be elucidated (reviewed in References 598 through 601). Proteins appear to vary widely in their tolerance to amino acid changes. Zeins, the prolamin storage proteins of maize, lack the essential amino acids lysine and tryptophan. In contrast to the failed attempts to modify phaseolin and CAT genes (noted above), modifications to a zein gene to introduce lysine and tryptophan codons (at several positions) do not affect the stability of the zein protein when the mRNAs are translated in *Xenopus* oocytes.[602] This result is promising in relation to the possibility of creating high-lysine corn by genetic engineering; stability of modified zein proteins in transgenic host plants now requires testing. Presumably, this protein can tolerate even severe modifications; a gross alteration (in which a 450 base pair open reading frame from a simian virus 40 coat protein was inserted into the coding region) did not affect zein protein stability, although its ability to aggregate and form protein bodies was affected.[602]

In general, seed storage proteins are significantly more stable in the seed of a foreign plant host than in the leaf, a factor that has impeded attempts to achieve high levels of protein accumulation in vegetative organs.[156,157,603,604] The low levels of accumulation of seed storage proteins observed in the transgenic host leaves, when compared with the seed,[585,603–605] may result from the continuous breakdown of the proteins in the protease-rich vacuoles of the leaves.[156] For example, despite similar levels of β-conglycinin mRNA in the leaves and seeds of transgenic petunia (utilizing a constitutive viral promoter to drive expression), the leaf cell environment yields a much lower level of protein accumulation.[585,604] A similar phenomenon occurs when pea vicilin is expressed in the leaves and seeds of transgenic tobacco.[156,584,605] Much higher levels of vicilin are detected in younger tissues (leaves) than in older ones, which is indicative that the rate of breakdown exceeds the rate of synthesis in these older tissues.[605] In many cases, it appears that the vacuolar targeting signals of seed storage proteins are recognized and processed correctly in the leaf cell environment because correct vacuolar transport of the protein occurs with a high degree of fidelity; however, there are important exceptions. Proteolytic processing of storage proteins commonly occurs upon their arrival in the vacuole. Differential proteolytic processing of storage proteins (e.g., β-conglycinin, vicilin) yields polypeptides of variable length in seed vs. nonseed tissues.[585,604,605] Analyses of proteolytic processing of vicilin deletion mutant proteins also reveal significant differences in leaf vs. seed tissues of the tobacco host (Figure 29); presently, it is not clear whether this reflects different subcellular sites of accumulation in the two tissues or whether there are differences in the substrate specificities of resident endoproteases.[584] The relationship between the differential processing of storage proteins and

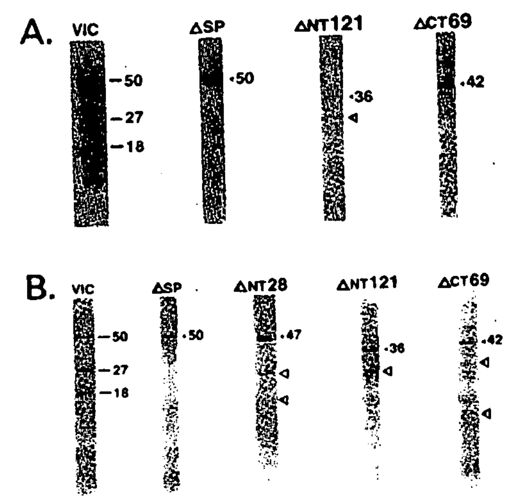


FIGURE 29. Proteolytic processing of pea vicilin deletion mutant proteins synthesized in the leaf and seed of transgenic tobacco. When expressed in transgenic tobacco seed, unmodified pea vicilin undergoes partial proteolytic processing yielding major polypeptide products of 50 (unprocessed), 27, and 18 kDa, indicative of a preferential usage of one of the two cleavage sites (P_1). In-frame deletions were made within the region of the vicilin gene, encoding mature protein, to eliminate the N-terminal 28 and 121 amino acids and the C-terminal 69 residues, while maintaining an intact signal peptide (ΔNT28, ΔNT121, and ΔCT69, respectively). (A) Mutant proteins synthesized in transgenic leaves. (B) Mutant proteins synthesized in transgenic seed. Open arrows indicate processed polypepides. All deletion mutant proteins undergo some normal proteolytic processing in the seed; the C-terminal deletion mutant remains unprocessed in the leaf. A modified vicilin gene, in which the DNA sequence corresponding to the signal peptide was removed (ΔSP), results in a polypeptide of 50 kDa in tobacco leaf and seed; none of the normal proteolytic cleavage products characteristic of expression of an unmodified vicilin gene are obtained (VIC: unmodified or intact vicilin protein synthesized in leaf or seed). (From Kermode, A. R., Fisher, S. A., Polishchuk, E., Wandelt, C., Spencer, D., and Higgins, T. J. V., *Planta*, 197, 501, 1995. With permission from Springer-Verlag, New York.)

protein instability due to proteolytic breakdown may need to be understood before the problem of low protein amounts in non-seed tissues can be fully addressed.

Because an instability of storage proteins in the leaf may, in part, be due to their (correct) vacuolar localization, one novel approach to overcome this problem is to "mistarget" proteins into subcellular compartments other than the vacuole.[156,157] Removal of the signal peptide sequence of a protein should abolish its transport to the leaf cell vacuole; entry into the ER lumen, and hence, the secretory pathway, is denied. Indeed, vicilin accumulation in the leaves of transgenic tobacco is increased by about fivefold when the DNA sequence corresponding to the signal peptide sequence is removed.[157] This modified vicilin gene results in a polypeptide of 50 kDa; none of the normal proteolytic processing products (characteristic of expression of an unmodified vicilin gene) are obtained, intimating that vacuolar transport is abolished. The fivefold increase in protein level is obtained from a level of vicilin mRNA that is lower than that obtained with the unmodified gene. Thus, the vicilin protein appears to be more stable as a consequence of its cytosolic localization.

Expression of a vicilin gene modified to encode an ER-retention signal (SEKDEL) at the carboxy-terminus of the protein results in a 100-fold increase in vicilin accumulation in the leaves of transgenic tobacco and alfalfa when compared with its unmodified counterpart.[156,157] This increase in protein is obtained without any change in vicilin mRNA level. Enhanced stability of the modified protein (which achieved a level of 2.5% of the total tobacco soluble leaf protein) may occur as a consequence of its new subcellular localization (i.e., predominantly in the ER), where, presumably, it receives some protection from proteolysis. Alternatively, there may be structural changes to vicilin as a consequence of its new carboxy-terminus that somehow render the protein more stable. Thus, targeting to a new subcellular locale may not be the sole factor involved in the enhanced protein stability. Manipulating the subcellular targeting of pea albumin 1 to increase accumulation of the protein in leaves of transgenic white clover and tobacco indicates that the endomembrane system is a relatively stable environment compared with the cytoplasm or chloroplast.[606]

The targeting and stability of engineered zeins in transgenic yeast cells have been examined.[607] Unmodified zein is targeted to (and accumulates within) the ER lumen in transformed yeast cells, as predicted for this maize storage protein. However, a truncated zein protein (in which the signal peptide and the first 36 amino acids of the mature protein are deleted) is synthesized in the cytoplasm and massively accumulates in the mitochondria, where it aggregates in protein body-like structures, similar to those present in maize endosperm cells. Thus, there appear to be distinct or separate domains of the zein polypeptide that, on the one hand, are responsible for membrane targeting; others may be responsible for the stability and aggregation of this storage protein.[607]

B. Enhancing Posttranslational Processes: Folding and Oligomer Assembly

The design of polypeptide sequences with a functional and stable three-dimensional conformation is impeded by our limited knowledge of the rules that govern protein folding and oligomer assembly and how these processes relate to a protein's ultimate stability in cells. Strategies must be developed to predict and avoid degradation of recombinant proteins in heterologous host cells, particularly in cases where these proteins contain sequences that render them sensitive to proteases, specifically targeting them for degradation.[608,609] A novel way to overcome some of the fundamental problems associated with protein design may be the construction of proteins *de novo* with unnatural (but highly stable) chain architectures by making use of the tools of synthetic chemistry (reviewed in Reference 610). It may also be possible to design modified storage proteins of desirable composition (e.g., rich in the sulfur-containing amino acid cysteine) that are stable and undergo correct folding and assembly in a manner reminiscent of natural (unmodified) seed storage proteins. Multiple cystine residues in a newly designed protein molecule may play an important role in stabilizing the protein; these have been shown to significantly stabilize the native structure of a protein,[611] and may also render proteins resistant to proteolysis.[612] Transient associations of seed storage proteins with molecular chaperones resident in the ER may also be a prerequisite for their correct folding and oligomeric assembly in foreign host cells. This may be achieved by ER-resident proteins that are endogenous in the foreign host cells, or it may require the introduction of genes encoding such "ER-helper" proteins from the donor genome. Glycosylation may play a general role in protein stability, primarily because of its promotive effect on protein folding; increasing the number of glycosylation sites (and their strategic positioning within a designer protein, for example, by site-directed mutagenesis) may increase both transport efficiency and stability.

A major system for selective protein degradation is the ubiquitin pathway in which proteins are committed to degradation by their ligation to ubiquitin as a multiubiquitin chain (reviewed in References 613 and 614). Although not yet tested in plant host systems, ubiquitin may be useful as a means of stabilizing proteins within heterologous hosts.[614–616] Interestingly, fusion of ubiquitin (by recombinant DNA techniques) to the amino-terminus of a protein can prevent it from degradation in yeast,[617] and results in increased accumulation, up to several hundredfold. The precise mechanism(s) responsible for such a dramatic increase in expression level is not clear. Ubiquitin may protect the amino-terminus of the fusion protein from proteolysis; alternatively, it may somehow guide the protein into a new compartment where it receives greater protection from proteases. Enhanced translation of the fusion protein may also have contributed to the increased expression. It will be interesting to determine whether a similar fusion of the ubiquitin gene to the coding region of a storage protein gene yields increased stability of the product in heterologous plant host cells.

VII. TARGETING OF PROTEINS TO THE NUCLEUS

A. Nuclear Structure and Components

Most of the DNA of eukaryotic cells is sequestered in the nucleus, an organelle delimited by two concentric membranes: the outer and inner nuclear membranes that together comprise the nuclear envelope (Figure 30). The outer nuclear membrane closely resembles the membrane of the rough ER, and like the latter, is studded with ribosomes engaged in protein synthesis. Proteins made on these ribosomes are transported into the space between the outer and inner nuclear membranes (the perinuclear space), which is continuous with the ER lumen.[94] The nucleus derives its mechanical strength from two networks of cytoskeletal elements; there is a thin meshwork of intermediate filaments lining the inner surface of the inner nuclear membrane (the nuclear lamina) and a less organized system of intermediate filaments surrounding the outer nuclear membrane. Plant cells also contain a nuclear lamina; its biochemical composition in pea nuclei appears to be similar to that in vertebrate and invertebrate animal cells, consisting of an array of 1 to 3 extrinsic membrane proteins, lamins A, B, and C and similar enzyme activities (nucleoside triphosphatase activity, derived from proteolysis from lamin A/C molecules in the nuclear scaffolding).[619,620] Specific proteins of the inner nuclear membrane act as binding sites for the nuclear lamina that supports it.

Proper functioning of the eukaryotic cell requires extensive bidirectional macromolecular traffic between the nucleoplasm and the cytoplasm. Nuclear import and export are highly specific processes, despite the dynamic nature of the nuclear envelope that dissolves during mitosis and is reconstructed after the completion of cell division, necessitating the reentry of nuclear proteins. The specificity of nuclear trafficking is due, in part, to the perforation of the nuclear envelope by nuclear pores (Figure 30). Each pore is formed by a large, elaborate structure known as the nuclear pore complex that traverses the envelope, bringing the lipid bilayers of the two membranes together around the margins of the pore. The complex is thought to be composed of more than 100 different proteins arranged with a striking octagonal symmetry; its estimated molecular mass is 125 million Da. Bidirectional transport is allowed through the pore complexes that contain one or more open aqueous channels, estimated by experimental means to have an effective size of 9 nm in diameter and 15 nm long — only a frac-

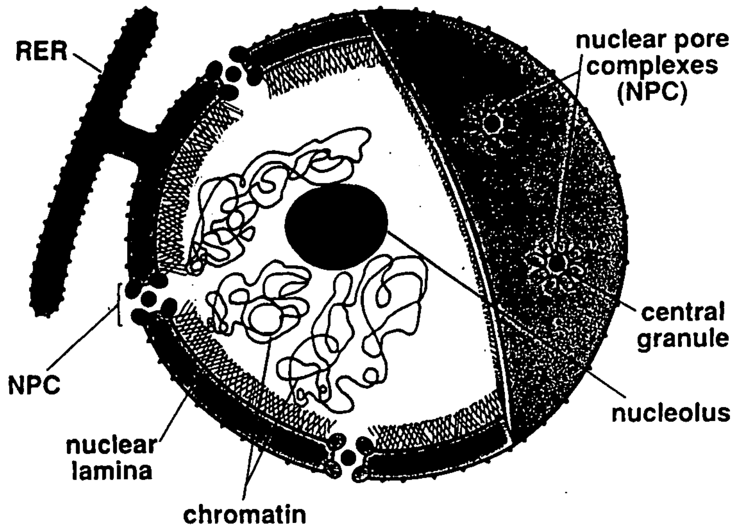


FIGURE 30. Schematic representation of the nucleus. (From Raikhel, N. V., *Plant Physiol.*, 100, 1627, 1992. With permission from the American Society of Plant Physiologists.)

tion of the total pore volume. Thus, the pore complex contains a pathway for free diffusion equivalent to a water-filled channel about 9 × 15 nm, such that molecules smaller than this size can passively diffuse (although exceptions exist). Because many cellular proteins are too large to pass by diffusion through the nuclear pores, the nuclear envelope allows the nuclear compartment and the cytosol to maintain different complements of proteins.

B. Mechanisms of Nuclear Targeting

1. Macromolecular Transport Across the Nuclear Envelope

Due to the presence of nuclear pores, transport into the nucleus is fundamentally different from that into other organelles, where proteins pass directly through the membrane. Included among large proteins selectively imported into the nuclear compartment from the cytosol where they are made are the histones, DNA and RNA polymerases, gene regulatory proteins, and RNA processing proteins. At the same time, tRNAs and mRNAs are synthesized in the nuclear compartment and then exported to the cytosol (Figure 31). Similar to the import process, the export process is highly selective; mRNAs, for example, are exported only after they have been correctly modified by RNA-processing reactions in the nucleus (e.g., by capping, polyadenylation, and splicing). Although many proteins remain within the nucleus after their import, others undergo shuttling (i.e., are continuously exported from the nucleus to the cytoplasm and are reimported into the nucleus). Although the significance of the shuttling process is not entirely clear, proteins that undergo this event have been identified in a range of eukaryotic cell types and include nucleolar proteins, steroid hormone receptors, heat-shock proteins, heterogeneous nuclear ribonucleoprotein (hnRNP) proteins, and the U1 small nuclear RNP (snRNP)-specific protein U1A. In some cases, the transport process is complex: ribosomal proteins, for example, are made in the cytosol, imported into the nucleus — where they assemble with newly made ribosomal RNA into particles — and then are exported again to the cytosol as part of a ribosomal subunit; each of these steps involves selective transport across the nuclear envelope.[94] In some cases, nuclear import of proteins must be strictly regulated by developmental or environmental cues, as in the case of gene regulatory proteins, which exert their control only at certain times or only under certain conditions. Plants, which must be more adaptive than animals to cope with changing environmental conditions, have likely evolved highly efficient regulatory mechanisms.[618]

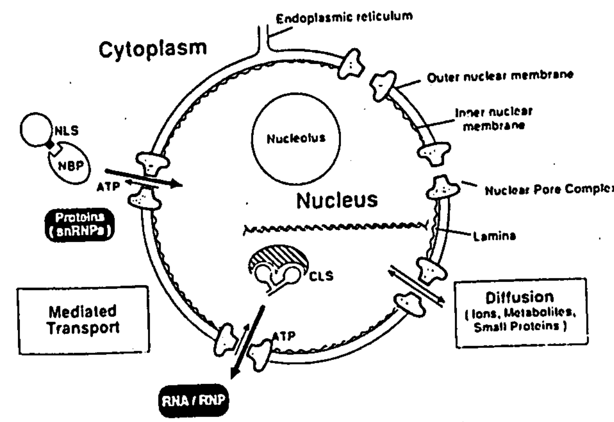


FIGURE 31. Nuclear import and export. See text for details. NLS, nuclear localization signal; NBP, NLS-binding protein; CLS, cytoplasmic localization signal. (From Nigg, E. A., Baeuerle, P. A., and Lührmann, R., *Cell*, 66, 15, 1991. With permission from Cell Press.)

2. Protein Import in Animal and Yeast Cells

When proteins are extracted from the nucleus and microinjected back into the cytosol of animal cells, even the very large ones are efficiently accumulated in the nucleus. The selectivity of nuclear import of proteins in all eukaryotic cells is due to the presence of a nuclear localization signal (NLS) that is found only in nuclear proteins (reviewed in References 618 and 621 through 629). Various approaches have been used to examine import of specific proteins through nuclear pores and to define the NLSs on these proteins. The role of nuclear pores in the signal-mediated transport of nucleoplasmin (a large protein consisting of distinct head and tail domains) was demonstrated by electron microscopy using the nucleoplasmin tail (which contains the NLS) coupled to spheres of colloidal gold. Following microinjection into the cytosol of *Xenopus* oocytes, the attached nucleoplasmin tails direct entry of the gold particles into the nucleus via the nuclear pores.[630] Defining the NLSs on proteins usually involves examining the fate of altered proteins in which mutations or deletions are made in a putative localization motif; this allows a determination of the sequences necessary for nuclear import. To determine whether the signal is sufficient for nuclear import, the putative NLS is linked to a cytosolic reporter protein.

Nuclear transport can be resolved into two steps: energy-independent targeting followed by ATP-dependent translocation, which is rate-limiting (reviewed in Reference 625). These two phases of import to the nucleoplasm are mediated by at least two functionally and biochemically distinct classes of NLSs.[622,628] NLSs were first identified in the large viral protein called T-antigen, which is encoded by SV40 (the Simian virus 40), and is needed for viral DNA replication in the host cell nucleus.[631,632] Many eukaryotic nuclear proteins carry T-antigen-like NLSs; these are related to the single motif of the SV40 large T-antigen NLS, comprised of a short stretch of basic amino acids (**PKKKRKV**) or the bipartite nuclear localization motif of nucleoplasmin — two regions of basic amino acids separated by a spacer of more than four residues (**SPPKAVKRPAATKKAGQAKKKKLDKEDES**).[633] Mating type (Mat) α2-like NLSs consist of a short hydrophobic region that contains one or more basic amino acids (**KIPIK**); in contrast, the NLSs of some viral nuclear proteins are quite different (e.g., AAFEDLRVRS is the NLS of the influenza ribonucleoproteins). However, widely distinct from all these signals are those comprised of a trimethylguanosine cap that have been found in U snRNPs and at least one U snRNA-binding protein.[634]

Nuclear import is thought to involve at least two parallel nuclear localization pathways whereby targeting to the nuclear pore complex is mediated by a separate apparatus.[625] Proteins bearing T-antigen-like NLSs compete for a common receptor; the nuclear pore complex-mediated import of U1, U2, U4, and U5 snRNPs occurs by a kinetically distinct targeting pathway (see Reference 635 and references therein).

In vitro studies in which nuclear import is reconstituted indicate that isolated nuclei are not sufficient to support nuclear import; cytosolic factors including NLS binding proteins and chaperones are required. An *in vitro* nuclear transport assay,[636] in which cytosolic extracts from *Xenopus* oocytes were added to digitonin-permealized cultured BRL cells, was used to fractionate and reconstitute the import process.[637] A pivotal finding was that separate cytosolic factors act in concert to mediate both the targeting and translocation phases of nuclear import. One fraction (an NEM-sensitive fraction, termed A) is thought to contain the targeting apparatus and to mediate the association of the protein with the nuclear envelope; the second fraction (B) was derived from a whole-cell extract containing soluble components from both the cytosol and nucleus, and promotes the ATP-dependent translocation of nuclear envelope-bound proteins into nuclei (reviewed in Reference 625).

Biochemical, functional, and genetic approaches have been employed in studies to isolate nuclear import factors and elucidate their specific functions (e.g., as putative receptors for signal recognition) and import factors that stimulate translocation (reviewed in Reference 624). Import receptors are expected to recognize NLSs, stimulate nuclear translocation, and interact with the nuclear pore complex. NLS-binding proteins have been identified by various means, including NLS affinity chromatography.[638] Only partial fulfillment of these properties is satisfied by putative receptors isolated so far. Purified binding proteins from anucleate bovine erythrocytes (putative NLS targeting receptors) are able to stimulate nuclear import in permealized cells;[639] however, a further finding is that multiple NEM-sensitive cytosolic factors are required for nuclear import, one of which has an NLS binding activity. More recently, reconstitution of the first step of nuclear transport (binding to the pore complex), occurred in permealized animal cells using two purified components — a previously identified NLS receptor and a 97-kDa protein of bovine erythrocytes.[640] Phosphorylation-dependent NLS-binding proteins (NSP70s) are also likely candidates for NLS receptors, and have been identified in a number of eukaryotes, including yeast, fruit flies, human cell cultures, and corn.[641] These proteins may be highly conserved because antibodies against yeast NSP70 inhibit *in vitro* protein import into permealized *Drosophila* cells. Further studies will be needed to systematically determine which proteins having NLS-binding activity are functional in nuclear import assays; and conversely, which of the functionally defined NEM-sensitive import factors and genetically defined import factors also demonstrate NLS-binding activity. Also requiring investigation are the mechanisms by which receptors interact with the nuclear pore complex; for example, whether any pore proteins mediate this event.[624] The available evidence suggests that NLS-binding proteins are able to recognize diverse NLSs, and that only a few types of import receptors are required for protein import. Also being characterized are the integral proteins of the nuclear pore complex that may have roles in the anchoring of the complex in the nuclear pore and regulation of nucleocytopathic traffic through the NPC.[642,643]

Chaperone proteins of the stress-70 families (Hsp70 and Hsc70) function in nuclear import,[644] although it is not clear yet whether their role is a direct or indirect one. For example, it has not been shown whether they interact directly with proteins destined for the nucleus or with a component of the translocation apparatus.[625] Stress-70 proteins function during transport into the ER lumen (see earlier) to help maintain nascent precursors in an extended conformation required for translocation competence; a similar facilitory process occurs during translocation of proteins into other organelles (e.g., chloroplasts and mitochondria). However, unlike proteins translocated into these organelles, proteins are generally not imported into nuclei as extended polypeptide chains.[622,625,628] Multisubunit proteins and RNPs with diameters in excess of 20 nm (the approximate dimensions of a 500-kDa spherical protein) are imported; karyophilic gold particles of 26 nm can also be accommodated by the translocation apparatus. What, then, is the specific role of these chaperone proteins in nuclear import? One suggestion is that stress-70 proteins bind to the surface of folded proteins destined for import and affect the exposure or conformation of the NLS;[625] indeed, the context of the NLS has been demonstrated to be important in the efficiency of nuclear import (reviewed in Reference 618). Another suggestion is that they may associate with a component of the cytosolic or nuclear pore complex-associated apparatus. A role in the disassembly of a targeting complex with the nuclear pore complex represents another possibility, somewhat similar to the role of the uncoating ATPase of clathrin-coated endocytic vesicles.[368] Finally, it cannot be ruled out that the stress-70 proteins may serve as the ATP-hydrolyzing motor that catalyses vectorial translocation through the nuclear pore complex-associated cytoplasmic filaments.[622]

As mentioned earlier, in all eukaryotic cells, there must be developmental and environmental controls over the nuclear import of certain proteins. Thus, although NLSs often function constitutively, their activity can also be regulated by phosphorylation of amino acids that are a component of, or flank, the NLS, or by distant regions of the polypeptide chain that can either mask the signal or anchor the protein in the cytoplasm (reviewed in Reference 645; see references therein). A striking example of regulated nuclear import and its role in the control of gene expression is the dorsal protein of *Drosophila*. This protein is the morphogen that determines the dorsoventral polarity of the embryo. It is found in the cytoplasm during early development; however, during cleavage it becomes localized to the nucleus exclusively in the ventral portion of the embryo, where it activates the transcription of the ventralizing genes, *twist* and *snail*. Aberrant development occurs when this regulation of nuclear transport fails. There are also examples of proteins with both constitutive and regulated NLSs; one such protein is the nuclear oncoprotein encoded by the adenovirus type 5 *Ela* gene. The constitutive NLS is of the SV-40 type and is located in the C-terminus, whereas the regulated NLS has a nonconventional motif and resides within amino acids 140–185. The latter can direct nuclear import in injected *Xenopus* oocytes, but not in transfected cells, providing evidence for its regulated nature. Nuclear import via the regulated NLS is ATP-dependent and competes with the C-terminus NLS. In DNA-injected fertilized *Xenopus* eggs, the NLS directs nuclear import up until the early neurula stage in all embryonic tissues; subsequently, the use of the NLS is turned off in specific tissues at specific times. Thus, there appears to be a hierarchy among the embryonic germ layers as to when the second signal becomes nonfunctional.[645]

3. Nuclear Import and Export of Proteins that Undergo Shuttling

As mentioned earlier, some proteins continuously shuttle between the nucleus and the cytoplasm. Some of the proteins may be transported via substrates to which they are bound and disso-

ciate only after crossing the nuclear envelope. For example, certain nuclear mRNP proteins associated with mRNA transport, which do not necessarily function directly in transport, are shed only after the mRNP is exported to the cytoplasm. These mRNPs are subsequently recycled back to the nucleus.[646] Shuttling may also be a common property of signal receptors and chaperones in nuclear transport. Oocyte stress-70 proteins, for example, shuttle between the cytoplasm and the nucleus;[647] likewise, binding proteins that recognize NLSs on nucleolar proteins and unique components participating in the export of ribosomal subunits to the cytoplasm are likely in dynamic equilibrium across the nuclear envelope, undergoing frequent shuttling. The nucleolar transport factor, Nopp140 from rat liver, binds NLS sequences and is localized predominantly to nucleoli but can be found in the nucleoplasm and transiently in the cytoplasm, even in the absence of protein synthesis.[648] Shuttling is suggested to occur along linear tracks that extend from sheet-like structures in the nucleolar interior out through the nuclear pore complexes into the cytoplasm.[648] In addition to mediating carrier-type functions, a further role of the shuttling process may be to mediate feedback control on the synthesis of nuclear proteins at the levels of translational control or mRNA degradation.[621]

It is now clearly established that shuttling does not require specific export sequences and that it can be decreased by sequences that promote binding in the nucleus (reviewed in Reference 621). The relative contributions of selective retention and selective entry to protein accumulation in the nucleus have also been examined. The view most supported by the available evidence is that, although import into the nucleus is a selective process, accumulation can be reinforced by selective binding once a protein has been transported into the nucleus.[649] Without this selective binding, some nuclear proteins will escape and shuttle repeatedly between the nucleus and the cytoplasm. Clearly, there are proteins of the nucleus that do not exhibit shuttling activity, but rather, are stably retained within the nucleus. One such protein is lamin B2, which comprises part of the nuclear lamina; removal of sequences required for its incorporation into the lamina converts it from a nonshuttling to a shuttling protein.[621] The question of whether export of shuttling proteins is a passive or active process has also been addressed, but has yielded conflicting results. However, it is clear that exported mRNAs and ribosomal subunits leave the nucleus much faster than the smaller shuttling proteins, which further supports the view that RNP export is a facilitated, energy-dependent process.

4. Nuclear Transport In Plants

Recent studies have yielded some information on the requirements for nuclear import in plant cells.[618,650-667] As in studies of other eukaryotic organisms, the two criteria used to define the NLSs of plant nuclear proteins are (1) sequences necessary for correct localization of the protein within the nucleus; deletion or mutation of certain key sequences within the NLS results in a loss of nuclear transport or inefficient transport leading to localization in both the cytoplasm and the nucleus; and (2) sequences sufficient to redirect a cytoplasmic reporter protein to the nucleus.[618] Transient expression systems using tobacco protoplasts or onion epidermal layers have been particularly useful.

Efforts have been made to determine whether the localization signals of other eukaryotic nuclear proteins are recognized in plant cells.[654,657] As mentioned earlier, the SV40 large T-antigen contains a short stretch of basic amino acids (**PKKKRKV**) that serves as a signal for entry of the viral protein into the animal cell nucleus and is also found in a number of other eukaryotic nuclear proteins. This nuclear localization motif is also recognized in plant cells. Transient expression of a T7 RNA polymerase gene in tobacco protoplasts results in nuclear localization of the corresponding protein, provided the gene construct is modified to contain sequences encoding the NLS of the SV40 large T-antigen.[654] Likewise, a chimeric protein, comprised of the 7 amino acid T-antigen NLS fused to the cytosolic GUS protein, localizes to the nucleus in transformed tobacco cells.[657] A mutant form of the NLS, in which an essential K residue is mutated to T, abolishes nuclear transport of the GUS protein. Thus, at least some aspects of nuclear transport are common between plant and animal cells.

a. Nuclear Localization of Gene Regulatory Proteins

The nuclear import mechanisms of two gene regulatory proteins from maize have been analyzed.[618,658-660,667] Opaque-2 is a *trans*-acting factor that regulates the expression of the 22-kDa zein genes of maize;[668] it belongs to the basic domain, leucine-zipper (bZIP) class of proteins. It localizes to the nucleus in maize endosperm tissue and, when expressed in transgenic tobacco, the fidelity of nuclear transport is preserved.[658] Further, a chimeric protein comprised of Opaque-2 linked to the normally cytosolic GUS, is correctly localized to the nucleus in both transgenic tobacco and transiently transformed onion cells.[659] Two independent domains of the Opaque-2 protein are sufficient to redirect the GUS reporter protein to the nucleus: (1) NLS A, located between the two transcriptional-activating domains, that has a SV40-type motif and is less efficient in directing GUS to the nucleus; and (2) NLS B, located in the highly conserved DNA-binding domain, having a bipartite structure and characterized by being more efficient in directing nuclear transport of GUS. To date, studies of two other bZIP proteins from plants (TGA-1a and TGA-1b) and three bZIP proteins from animals (Fos, Jun, and Eb1) have localized the NLSs within the basic domain associated with DNA binding.[657,666,669,670] Both NLS A and NLS B are necessary for nuclear targeting of Opaque-2.[660] Further analysis of the NLS B region of Opaque-2 to identify the critical amino acid residues reveals that both the bipartite structure and the net basic amino acid content are important for nuclear targeting. A mutation that contains altered amino acids on both parts of the bipartite NLS severely reduces nuclear targeting; this led to the definition of two classes of bipartite NLSs, distinguished by the presence or absence of acidic amino acids in the spacer region between the two basic motifs. Analysis of a mutant form, in which a conservative mutation is made within NLS B, demonstrates that the nuclear targeting function of this domain is independent of DNA-binding; while completely abolishing DNA binding, the mutation leaves nuclear targeting unaffected. The bifunctionality of this domain may be conserved in all members of the leucine zipper class; to attain DNA-binding specificity, basic amino acids may require a specific arrangement in the proper configuration for NLS function.[660] Biochemical studies of transport into isolated plant nuclei using the Opaque-2 bipartite NLS (and various mutant forms) may aid in identifying putative receptor proteins in plant cells.[661]

Another transcription-activating protein from maize (R protein) contains a helix-loop-helix motif similar to those found within a number of animal transcriptional activators (e.g., myo D1 and myc).[671] Fusion proteins in which the R protein is linked to GUS are successfully targeted to the nucleus in onion epidermal cells, regardless of whether the R protein portion of the chimeric protein is at the N-terminus or C-terminus (i.e., R-GUS or GUS-R, respectively).[667] Further analysis of chimeric constructs containing different regions of the R gene fused to the GUS gene coding region led to the identification of three specific NLSs capable of redirecting the GUS protein to the nucleus. The first of these is a 10 amino acid N-terminal NLS-A, comprised of several arginine residues; a similar localization signal is present in only a few viral proteins. The medial NLS-M (also comprised of 10 amino acids) contains an SV-40 motif, and the carboxy-terminal NLS-C is a MAT α-2 type. NLS M and C are independently sufficient for nuclear transport when fused to the amino-terminus of GUS; these two nonhomologous signals may be involved in different steps of nuclear transport or may interact with different import components of similar function.[618] NLS A fused to GUS leads to partitioning of the chimeric protein between the nucleus and cytoplasm. The position of the NLS in the transported protein is important; partitioning between the nucleus and cytoplasm also occurs when all three NLSs are independently fused to the C-terminal portion of GUS.[667] Deletion analysis of the three specific NLSs reveals that multiple signals are necessary for nuclear targeting of the R protein. Only the combined action of NLS A and M, or NLS C and M, are able to direct nuclear localiza-

tion. GUS fusions comprised of all other combinations become localized in both the nucleus and the cytoplasm.

b. Nuclear Localization of Plant Virus Proteins

Plant viruses often encode a number of different proteins that accumulate in different subcellular compartments after infection. Members of the potyvirus group form characteristic inclusions after infection, including nuclear inclusions composed of two proteins involved in replication of potyviral RNA, NIa and NIb.[655] The role of nuclear transport of the two proteins is somewhat puzzling because RNA replication apparently occurs within the cytoplasm. Fusion of either NIa or NIb to GUS is sufficient to direct GUS reporter activity to the nucleus in transfected protoplasts and stably transformed tobacco plants. Deletion analysis of the NIa gene identified a bipartite NLS within the protein consisting of two short regions (comprised of 11 and 30 amino acids, separated by 32 residues) located within the amino-terminal region of the protein.[650] Deletion and mutational analysis of NIb yields a somewhat different scenario.[665] Amino acid changes affecting two areas, between residues 3–5 and 303–306, abolished nuclear transport. However, the assignment of NLS function to these regions is complicated by the finding that substitutions at four additional sites throughout the NIb sequence also render fusion proteins primarily cytoplasmic. Further, each of six deletions in NIb debilitates nuclear localization, regardless of whether the basic clusters are deleted. Insertion of Pro-Pro dipeptides, predicted to lead to abnormal folding of NIb, reduces nuclear import when placed at three out of four positions. Thus, nuclear localization of this protein may require a stringent tertiary structure in addition to one or more NLSs.[665]

c. Nuclear Localization of Nucleic Acids during Agrobacterium*-Mediated Plant Transformation*

Agrobacterium tumefaciens is a soil pathogen capable of infecting a number of plant species (primarily dicots) and inducing crown gall disease. Tumorous growths (crown galls, characterized by continuous cell division) arise on these plants at wound sites as a consequence of the production of growth hormones (auxin and cytokinin) in infected and transformed plant cells. A large plasmid within the bacterium (the Ti plasmid) effects the transfer of tumor-inducing genes (genes encoding enzymes required for the biosynthesis of the hormones) as a single-stranded DNA molecule to the plant nuclear genome (reviewed in Reference 672). The T-DNA of *Agrobacterium* is transferred to the plant cell and eventually to the plant nucleus as a complex of three components: a single-stranded DNA molecule, the T-strand, and two different virulence proteins — vir D and vir E. The vir D2 protein attaches covalently to the 5′ end of the T-strand; over 600 copies of the vir E2 protein bind along the length of the T-strand, making the entire T-complex approximately 60 times longer than the diameter of the nuclear pore.[651] The vir D2 protein that is tightly attached to the T-DNA is thought to pilot the T-strand to the plant nucleus. A bipartite NLS at the carboxy-terminus of vir D2 is sufficient for nuclear import of GUS; it also appears to be sufficient for efficient T-DNA transfer to the plant cell nucleus.[652,653,662–664] The vir E2 protein may also play an important role in nuclear transport of the T-strand.[651] When vir E2 is fused to GUS, the chimeric protein localizes to the nucleus in tobacco cells; a role *in planta* is suggested by the finding that tumorgenicity of an avirulent vir E2 mutant is restored when inoculated on transgenic plants expressing the vir E2 protein. Nuclear localization of vir E2 is mediated by two bipartite NLSs (NSE1 and NSE2), and efficient nuclear localization of GUS is achieved only when both signals are present. An interesting suggestion is that vir E2 acts as a molecular chaperone serving to coat, unfold, and target the T-strand to the nucleus. Further studies to elucidate the cooperation between the two virulence proteins in effecting nuclear transport of the T-DNA are awaited with interest.

VIII. PROTEIN TARGETING INTO PEROXISOMES AND GLYOXYSOMES

Peroxisomes (microbodies) are thought to be present in all eukaryotic cells; they share a num-

ber of common characteristics, as well as some unique ones, depending on the organism and developmental stage. All of these organelles, regardless of their origin, are bounded by a single membrane, do not possess an organelle genome, and contain catalase for the breakdown of hydrogen peroxide (reviewed in References 673 and 674). Although they do not contain any internal membrane structure, the matrix of the organelle occasionally contains crystalline or fibrillar inclusions that contain enzymes. Their constituent proteins are encoded by nuclear genes, are synthesized on free polyribosomes, and are imported from the cytosol. Higher plants contain several classes of peroxisomes that carry out different metabolic roles; these are found during different stages of the plant lifecycle and may be species-, organ-, or cell-specific.[675] At least three classes of peroxisomes (microbodies) have been defined: (1) glyoxysomes, present in postgerminative seedlings and senescent organs, that function in the mobilization of seed storage lipids by housing the enzymes of the glyoxylate cycle;[676–678] (2) leaf-type peroxisomes, present in photosynthetically active tissue, that contain enzymes essential for the light-dependent reactions of photorespiration, including glycolate oxidase and hydroxypyruvate reductase;[677,679] and (3) specialized peroxisomes in root nodules of certain legumes, which contain urate oxidase involved in the production of ureides, the primary nitrogenous products exported from nodules. In addition, plant organs (e.g., roots) also contain unspecialized peroxisomes without defined metabolic roles.[677,680] Plant peroxisomes are functionally adaptable organelles; they can change their specialized metabolic roles (dictated by the unique set of enzymes that they accumulate) in response to the specific requirements of the cell that are, in turn, dependent on the specific developmental stage. For example, during seedling development, these organelles undergo an interconversion in function during the transition from heterotrophic to autotrophic growth.[681–685] During greening of cotyledons, glyoxysomes are converted to leaf peroxisomes, which function in photorespiration. During senescence of leaves and cotyledons, a reverse transition occurs and glyoxylate cycle enzymes are once again found in peroxisomes; enzymes characteristic of both pathways coexist within the same organelle at the stage of transition from peroxisome to glyoxysome.[681,683,684]

Insight into the mechanisms for targeting proteins to peroxisomes in plants has been derived by studies of protein import *in vivo* (e.g., examining the subcellular localization of peroxisomal proteins in transgenic hosts); *in vitro* import systems using isolated organelles have also been developed. A carboxy-terminal tripeptide conforming to the consensus sequence S/A/C-K/R/H-L (the so-called SKL motif) has been identified as a conserved peroxisomal targeting sequence.[686] The initial characterization of this signal was based on gene transfer experiments with the firefly luciferase gene,[687] which showed that the carboxy-terminal three amino acids (SKL) of the luciferase enzyme are both necessary and sufficient for routing of the protein into peroxisomes. Correct routing of firefly luciferase into the peroxisomes of diverse organisms reveals a conservation of the signal recognition process and import mechanism;[688] furthermore, the tripeptide motif is present at the carboxy-terminus of many peroxisomal enzymes from animals, plants, and yeast (reviewed in Reference 689). Despite the remarkable conservation of this C-terminal targeting signal, there is substantial evidence in all eukaryotic organisms that multiple signals are involved in the targeting of peroxisomal matrix proteins. Further, if a tripeptide motif is involved, it need not be located at the extreme carboxy-terminus of the protein. An internal SHL sequence occurs within the 27 amino acid C-terminal topogenic peptide of human catalase; likewise, a signal in rat thiolase is able to specify peroxisomal import, even though located internally (reviewed in References 689 and 690). The peroxisomal targeting signal of rat liver acyl-CoA oxidase resides in the carboxy-terminus of the protein; however, the same enzyme from *Candida tropicalis* contains two internally located regions consisting of about 100 amino acids that function in peroxisomal import. An amino-terminal-targeting signal has been identified in rat liver peroxisomal 3-ketoacyl-CoA thiolase.[690]

A survey of the carboxy-terminal sequences of plant peroxisomal proteins shows that many, but not all, contain a terminal tripeptide that is

similar to the luciferase consensus sequence (reviewed in Reference 673; see references therein) (Table 11). However, on some plant proteins, potential targeting determinants are also found upstream of the carboxy-terminus. For example, catalase and hydroxypyruvate reductase from cucumber lack the terminal tripeptide, but contain a similar tripeptide several amino acids upstream. Uricase and glycolate oxidase may contain two of these putative targeting sequences, separated by a single amino acid residue. The glyoxysomal enzymes isocitrate lyase and malate synthase contain a carboxy-terminal tripeptide similar to luciferase and other peroxisomal proteins. However, experiments to determine the function of the carboxy-terminal tripeptide on castor bean isocitrate lyase reveal that it is not essential for import into glyoxysomes in an *in vitro* system.[691]

The import of *in vitro*-synthesized isocitrate lyase into glyoxysomes isolated from sunflower cotyledons is temperature and ATP dependent. Progressive carboxy-terminal truncations of the isocitrate lyase gene were transcribed and translated to yield polypeptides with the same amino-terminus, but lacking in varying amounts of the carboxy-terminus. All of the truncated polypeptides were imported with the same characteristics as the full-length protein, suggesting that the targeting information resides within the first 168 amino acids. Thus, the carboxy-terminus containing the luciferase-like tripeptide is dispensable for targeting and import.[691] The putative C-terminal targeting signal on glycolate oxidase has also been tested for functionality in transgenic tobacco;[692] the last six C-terminal amino acids of the peroxisomal protein are sufficient to target cytosolic GUS to tobacco leaf peroxisomes. The carboxy-terminus of watermelon malate dehydrogenase does not contain an obvious targeting tripeptide, but this protein has an amino-terminal

TABLE 11
Carboxy-Terminal Sequences of Plant Peroxisomal Proteins

Malate synthase	
Brassica napus	L T L A V Y D H I V A H Y P I N * A **S R L**
Castor bean	L T L D A Y N N I V I H Y P K G * S **S R L**
Cucumber	L T L D A Y N Y I V I H H P R E * L **S K L**
Cottonseed	L T L D A Y N Y I V I H H P K D V S **S K L**
Isocitrate lyase	
Castor bean	T R P G A M E M G S A G S E V V A K **A R M**
Cotton	T R P G A G N I G S E G N L V V A K **A R M**
Brassica napus	T R P G A A G M G E G T S L V V A K **S R M**
Catalase	
Cucumber	A **S R M** N V R P N I
Cottonseed	S Q A D K S V G Q K L A **S L L** N V R P S I
Sweet potato	T Q A D R S L G Q K V A **S R L** N I R P T M
Maize	S Q C D A A L G Q K L P **S R L** N L K P S M
Glycolate oxidase	
Spinach	R S H I A A D W D G P * * S **S R A** V **A R L**
Lentil	R E H I V A D W D T P R I Q **P R A** L **P R L**
Uricase II	
Soybean	L P T D E P H G S I Q A S L **S R L** W **S K L**
Hydroxypyruvate reductase	
Cucumber	E N V S P P A A S P S I V N **A K A** L G N A
Malate dehydrogenase	
Watermelon	L E K A K K E L A G S I E K G V S F I R S

Note: Putative targeting sequences are indicated in bold type. The asterisks represent introduced spaces to better align the sequences.

From Olsen, L. J. and Harada, J. J., in *Molecular Approaches to Compartmentation and Metabolic Regulation*, Huang, A. H. C. and Taiz, L., Eds., American Society of Plant Physiologists, Rockville, MD, 1991, 129. With permission from the American Society of Plant Physiologists. See references therein.

extension that may play a role in transport.[693] This extension is comprised of 37 amino acids, has a net positive charge, lacks a long stretch of hydrophobic residues, and contains a cluster of serine residues; an AHL tripeptide within the extension may represent the necessary topogenic information. Notably, the N-terminal extension is cleaved off when associated with, or imported into, glyoxysomes. It has been suggested that at least two distinct import pathways exist for peroxisomal proteins that contain uncleavable C-terminal signals, versus those that contain peroxisomal import signals comprised of an N-terminal, cleavable presequence.[694] The precursor watermelon malate dehydrogenase is correctly targeted to and imported into peroxisomes of the methylotrophic yeast *Hansenula polymorpha*, indicating that the targeting signal of the plant protein is correctly recognized in this heterologous host. However, processing of the N-terminal sequence does not occur, suggesting that a specific peroxisomal peptidase is either absent or does not recognize processing signals within the heterologous protein.[694]

In addition to the diversity of signals that mediate the import of matrix proteins, there is evidence that peroxisomal membrane proteins use alternative mechanisms; these proteins do not contain motifs similar to matrix-directed proteins such as luciferase. The existence of multiple peroxisomal signals has been particularly important in elucidating the protein targeting defect in humans that exhibit Zellweger syndrome. Membrane ghosts prepared from peroxisomes of certain Zellweger patients import the thiolase precursor, but fail to import other proteins such as acyl-CoA oxidase and catalase. Thus, the cells from these patients are competent to import proteins containing only one type of peroxisomal targeting signal and are incapable of recognizing signals of the SKL-type.[695] This likely indicates that the proteins containing thiolase- or SKL-type targeting signals are recognized by different receptors.[686] An intriguing question to be addressed is whether these receptors interact with the same or independent translocation machinery to facilitate import.

To examine the participation of integral membrane proteins in the import of proteins into plant glyoxysomes, a radioactively labeled peptide having a sequence corresponding to the last 12 amino acids of rat acyl-CoA oxidase (D-Tyr-HKHLKPLQSKL) was used to detect a receptor capable of recognizing the SKL motif.[696] Binding of the radiolabeled peptide to alkali-stripped glyoxysomal membranes is saturable, and 80% of the binding can be replaced by unlabeled peptide or glyoxysomal matrix proteins. Protease treatment of the alkali-stripped glyoxysomal membranes lowers the number of high affinity sites and destroys all low affinity sites. Characterization of specific receptor proteins involved in peroxisomal import in plants is awaited with interest.

Further elucidation of the requirements for peroxisomal targeting and import have come from studies on animal proteins. In firefly luciferase, recognition of the tripeptide signal is context dependent[687] (reviewed in Reference 686). Linker insertions within the amino-terminal half of the enzyme as well as large fusions between cytosolic reporter proteins and luciferase produce chimeric polypeptides that fail to be transported into peroxisomes, even though the signal is present at the carboxy-terminus. Thus, the signal likely must be accessible in the folded protein. Because many peroxisomal proteins are synthesized at the mature size, few studies have addressed the involvement of molecular chaperones in peroxisomal targeting. However, recent findings support the role of chaperones in maintaining peroxisomal proteins in a conformation-competent conformation.[697] Import of catalase into peroxisomes is retarded when assayed in the presence of aminotriasol, a specific inhibitor of catalase activity. The drug is likely to hinder the unfolding of catalase preventing it from assuming a translocation-competent state; alternatively, it may mask the targeting signal.[698] Further, the translocation of peroxisomal proteins (e.g., acyl-CoA oxidase, isocitrate lyase) requires ATP hydrolysis,[691,699] which mediates chaperone-assisted translocation of proteins into other organelles (e.g., those undergoing translocation across the membranes of the ER, chloroplast, and mitochondrion). Interestingly, a 72-kDa protein that binds to clofibrate (an inducer of peroxisomal proliferation) is a member of the stress-70 protein family.[700]

A pertinent question addressed in plants is whether the metabolic role of a peroxisome is determined by the capacity of the organelle to recognize targeting signals of only a subset of proteins.[674,675] This was investigated in relation to

the developmental transition of peroxisomal function by analyzing the targeting of glyoxysomal proteins (characteristically synthesized during postgerminative growth of seedlings or during senescence) to leaf-type and root peroxisomes. When the glyoxysome-specific enzymes isocitrate lyase and malate synthase are synthesized in transgenic *Arabidopsis*, they are imported into leaf-type and root peroxisomes; furthermore, the same targeting determinant is recognized by different classes of the organelle. The ability of glyoxysomal proteins to be transported into several classes of peroxisomes using a common targeting determinant suggests that protein import does not play a regulatory role in determining peroxisomal metabolic function. Rather, the specific metabolic role of this organelle may be dictated primarily by controls over the synthesis/stability of its constituent proteins. These "upstream" controls are in turn specified by the differentiated state of the cells in which the organelles are found.[674]

IX. PROTEIN TARGETING INTO AND WITHIN CHLOROPLASTS

The chloroplast is a complex organelle that carries out a wide range of metabolic processes; in addition to housing the entire photosynthetic machinery, it is responsible for several vital biosyntheses (e.g., of fatty acids, phospholipids, and amino acids), the interconversion of carbohydrate intermediates, and the final steps in the assimilation of inorganic nitrogen and sulfate. It is a semiautonomous organelle, equipped with the machinery necessary for the transcription and translation of the limited number of proteins that are encoded by its own genome. Despite this synthetic capacity, the great majority of proteins required for chloroplast function are encoded by the nuclear genome; following their synthesis on free polyribosomes of the cytosol, they are imported into the chloroplast, where further intraorganellar sorting may take place. The complexity of the organelle at the metabolic level is mirrored by a corresponding complexity at the structural level. It consists of three distinct membrane systems (the outer and inner envelope membranes and the thylakoid membrane), which enclose three distinct soluble subcompartments (viz., the interenvelope membrane space, the stroma, and the thylakoid lumen). Thus, all six components comprise targets for protein transport (Figure 32).

Protein transport into the stroma and thylakoids has been analyzed in some detail (reviewed in References 701 through 706) and is summarized in the next sections; reference is also made to Figure 33. At present, little information is available on the mechanisms of protein targeting to the outer and inner envelope membranes or to the interenvelope membrane space.

A. Protein Targeting to the Stroma

Biogenesis and subsequent translocation of proteins across the chloroplast envelope membranes into the stromal compartment have been investigated in detail. The protein studied most extensively in this respect is the small subunit of ribulose bisphosphate carboxylase-oxygenase (SSU-Rubisco), a nuclear encoded protein. Following synthesis in the cytosol and subsequent translocation into the stroma, the SSU associates with its counterpart, the large subunit (LSU) (which is synthesized within the chloroplast) to form a large oligomeric complex comprised of eight subunits of each type.[707] The general principles derived from studies on the SSU-Rubisco appear to be valid for the transport of similarly directed proteins (e.g., ferredoxin).[701] In large part, analysis has been by *in vitro* import systems,[708] in which the transport of radiolabeled proteins (often constructed by recombinant genetic techniques and synthesized by *in vitro* transcription and translation) into purified intact chloroplasts is assessed. Verification of import can involve treatment with external proteases; membranes protect imported proteins from degradation, whereas external proteins remain susceptible. Isolated chloroplasts of pea stored in liquid nitrogen in the presence of dimethyl sulfoxide (65 to 70%) remain intact upon thawing and are fully functional in the import of precursor proteins; preserved thylakoids also exhibit near optimal activity for protein integration.[708]

Proteins that must cross the outer and inner envelope membranes (i.e., stromal proteins and other proteins that undergo further intraorganellar

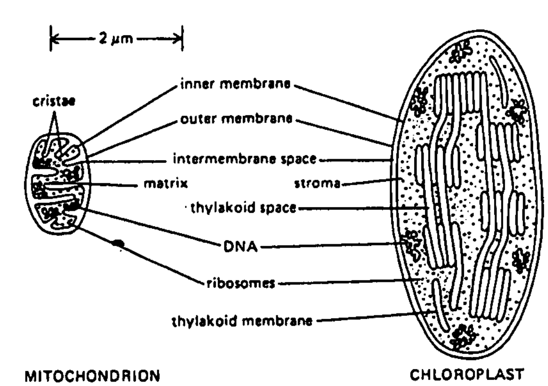


FIGURE 32. Schematic representation of a mitochondrion and a chloroplast. The chloroplast is generally much larger and contains a thylakoid membrane and thylakoid space. The mitochondrial inner membrane is folded into cristae. (From Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J., *Molecular Biology of the Cell*, 2nd ed., Garland Publishing, New York, 1989. With permission from Garland Publishing.)

targeting after reaching the stroma) are synthesized on cytosolic polyribosomes as precursors with a transient N-terminal sequence, the transit peptide. This transit peptide mediates both the initial binding of the precursor to the outer membrane of the chloroplast envelope as well as its posttranslational import. Coincident with translocation of the precursor, or shortly thereafter, a specific stromal peptidase effects transit peptide removal, yielding the mature protein. The proteolytic enzyme is chelator sensitive and is highly specific for imported precursors.[709] The precursor of SSU-Rubisco is processed to the mature size in two steps (involving the same enzyme, but different amino acid residues) and proceeds via an 18-kDa intermediate.[710,711]

The transit peptides of chloroplast proteins generally behave in an organelle-specific[712] and autonomous fashion; when attached to a heterologous (nonchloroplastic) passenger protein, (via gene fusion experiments, followed by *in vitro* transcription/translation), they are capable of effecting unidirectional import of the passenger protein into the chloroplast. Successful import into the stroma was first demonstrated with the transit peptide of SSU-Rubisco linked to bacterial reporter proteins (CAT and neomycin phosphotransferase II),[713,714] and subsequently has been shown with a number of different passenger proteins,[715-717] as well as with chimeric proteins having transit peptides from other chloroplast proteins.[701,712,718-720] Not all transit peptides may be equally efficient or competent in effecting translocation of a given passenger protein across the chloroplast membrane; likewise, different passenger proteins linked to the same transit peptide exhibit differences in the extent and efficiency of import.[702] Import of chimeric proteins can (in some cases) be increased when part of the mature chloroplast protein is added (in addition to the transit peptide), but this is not always a consistent result, and may, in part, be due to structural reasons (e.g., maintaining a particular secondary or tertiary configuration of the transit peptide that is conducive to import; see later discussion).[715,716,721,722] More systematic and quantitative studies are needed to evaluate these results.

Attempts to define the functional domains of transit peptides (e.g., those sequences involved directly in receptor binding or translocation) have

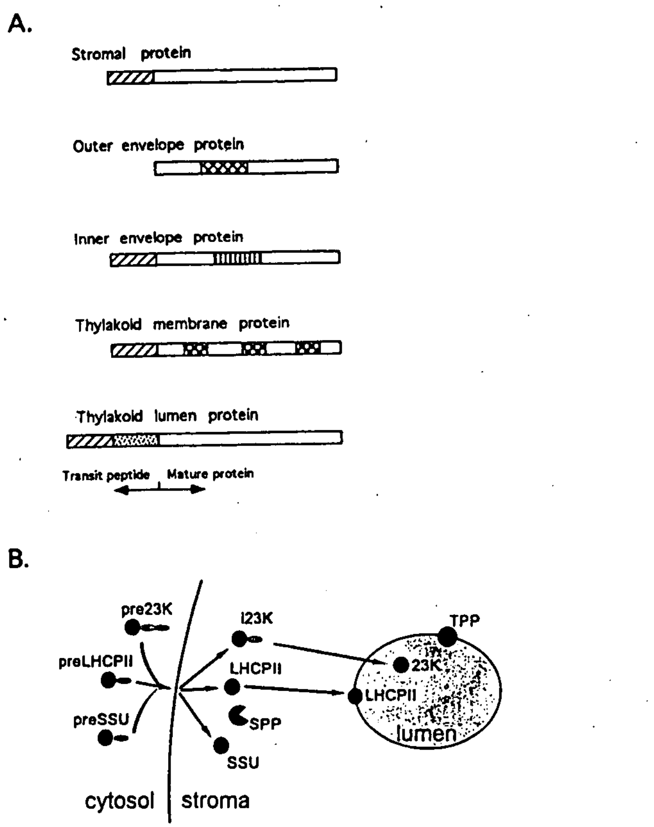


FIGURE 33. (A) Targeting information in chloroplast precursor proteins. In each protein, the mature portion is represented by the region indicated at the bottom of the figure. The right-leaning hatched bar represents an envelope transfer sequence; the spotted bar represents a thylakoid transfer sequence. Other topogenic sequences (cross-hatched bar, the outer envelope targeting domain; vertical-striped bar, the inner envelope targeting domain; checkerboard bar, the thylakoid membrane targeting signal patches) reside within the mature portion of the proteins and are poorly characterized; their relative positions are arbitrary. (From Theg, S. M. and Scott, S. V., *Trends Cell Biol.*, 3, 186, 1993. With permission from Elsevier Publishing, Cambridge.) (B) Basic pathways for the import of proteins into the chloroplast stroma, thylakoid membrane, and thylakoid lumen. The diagram depicts the basic import pathways for a stromal protein, Rubisco small subunit (SSU), a thylakoid membrane protein, LHCPII, and a thylakoid lumen protein, 23K. SSU and LHCPII are synthesized with stroma-targeting presequences (black ovals), which mediate binding to receptors on the chloroplast surface and transport across the envelope membranes. 23K is synthesized with a bipartite presequence containing a similar stroma-targeting signal followed by a thylakoid-transfer signal (hatched oval). The stroma-targeting signals of all three precursors are removed by the stromal processing peptidase (SPP). LHCPII inserts into the thylakoid membrane by means of signals in the mature protein, whereas the cleavable thylakoid transfer signal of 23K mediates translocation across the thylakoid membrane. Insertion of LHCPII requires ATP and a stromal protein factor; the requirements for thylakoidal protein translocation are shown in greater detail in Figure 34. (From Robinson, C. and Klösgen, R., *Plant Mol. Biol.*, 26, 15, 1994. With permission from Kluwer Academic Publishers.)

taken the form of two general approaches (reviewed in Reference 702). One approach has been the comparative analysis of the primary (and to a lesser extent, secondary) structures of transit peptides.[702,723] The other general approach has been to address structure-function relationships by generating and analyzing deletions (or specific changes to amino acid residues) in various regions of the transit peptide (e.g., those of ferredoxin and SSU-Rubisco).[721,724-728] With respect to the first approach, the amino acid sequences of the SSU-Rubisco transit peptides have now been determined from 48 genes representing 22 plant species.[702] From the statistical analyses of these and other transit peptides (e.g., those of other stromal and thylakoid membrane/lumen precursors), the general picture that emerges is that sequence similarities generally exist among transit peptides of the same precursor derived from different plant species. Few similarities are found among different precursors, even when the precursors are derived from the same plant species.[702] Despite the general lack of primary sequence similarities among transit peptides, some common features have emerged. They are rich in hydroxylated amino acids, serine and threonine, and also have a number of small hydrophobic amino acids such as valine and alanine. A general lack of acidic amino acids is evident; overall, they have a net positive charge.[702,723]

Small deletions in the transit peptide of SSU-Rubisco generally do not abolish import into chloroplasts.[724,725] Mutagenesis experiments (to effect changes in specific amino acid residues in various regions of the transit peptide) indicate that N-terminal and C-terminal sequences may be essential for binding and/or uptake.[701,725-727] However, the general concept emerging is that the essential features of transit peptides are not found in their amino acid sequence, but rather in some higher-order structure. Thus, specific conserved secondary or tertiary structural features in the transit peptide may be more important in signal recognition/decoding by receptors or other components of the import/translocation machinery.[701,702] The presequences of several mitochondrial proteins have the capacity to form amphiphilic structures (see later discussion), and amphiphilicity may be one essential feature of their function.[729,730] However, there are numerous differences between the topogenic signals of mitochondrial and chloroplastic proteins, such as absolute amino acid composition, length, and predicted secondary structure.[723] Furthermore, the presence of amphipathic β-sheet structures (predicted by hydrophobic moment analyses; see later discussion) has not been confirmed for chloroplastic peptides.[704,723] An alternative view is that chloroplast transit peptides are designed to be devoid of any regular secondary or tertiary structure.[731] If they are indeed 'perfect random coils,' a series of interactions with cytosolic and chloroplastic chaperones may be all that is required for chloroplast targeting and import.[731]

B. Protein Targeting Toward the Thylakoid Membrane System

The thylakoid membrane system encloses the thylakoid lumen and is itself completely surrounded by stroma. Thus, proteins destined for either subcompartment must undergo further intraorganellar targeting following their import into the stroma. Proteins of the thylakoid membrane or lumen utilize a targeting mechanism that consists of two independent steps; information is encoded for the sequential translocation across the outer and inner envelope membranes (envelope targeting signal) and for the integration into (or translocation across) the thylakoid membrane (Figure 33).[720,732-734] The most abundant thylakoid membrane protein imported from the cytosol is the light-harvesting chlorophyll binding protein (LHCP) of photosystem II. The transit peptide of this protein directs import of precursor into the stroma while the signal for subsequent insertion into the thylakoid membrane is encoded in the mature protein.[718,734,735] Stable integration into the thylakoid membrane may involve complex interactions between membrane-spanning domains and other domains in the mature protein, which have importance in protein refolding.[734,736] A soluble proteinaceous stromal factor is also required;[737] its primary function may be to maintain the solubility and integration competence of LHCP.[738] Proteolytic cleavage of LHCP to its mature size occurs only after membrane insertion is completed.[739] However, despite the lack of a thylakoid network in albinoid petunia protoplasts,

transient expression of the LHCP gene within this system yields a protein that is promptly processed into its mature-cleaved form.[740] The site of integration appears to be the exposed stromal (unstacked) thylakoid regions; from this site, LHCP diffuses to its final destination in the stacked granal membranes.[741] There is an energy requirement for integration of LHCP that can be met by ATP but is considerably more efficient when there is also a proton motive force (PMF) across the thylakoid membrane (see later discussion).[742]

The targeting and assembly of thylakoid membrane proteins are highly complex processes that appear to be specific and characteristic for each polypeptide species. For example, the apoprotein of the chlorophyll a/b antenna complex CP24 of photosystem II is a remote relative of the light-harvesting complex (LHC) apoproteins. However, the CP24 displays unique structural differences to LHC apoproteins that have a potential impact on the routing and targeting processes during biogenesis.[743] In particular, CP24 lacks a pronounced second hydrophobic segment present in the mature polypeptide chain of LHCPs, and carries a transit peptide that is reminiscent of thylakoid-targeting transit peptides (see subsequent discussion). Experiments using both radiolabeled authentic precursor and chimeric proteins demonstrate that the transit peptide of the CP24 apoprotein is required only for import of the protein into the organelle. All subsequent steps, such as integration of the protein into the thylakoid membrane, binding of chlorophyll, assembly into the CP24 complex, and migration to the grana lamellae still take place if the authentic transit peptide is replaced by a targeting signal of a nuclear-encoded stromal protein.[743]

Nuclear-encoded thylakoid lumen proteins must transit all three chloroplast membranes (viz., the outer and inner membranes of the envelope and the thylakoid membrane) and hence, have a uniquely complex import pathway (reviewed in References 703 and 706). Domain swapping and fusion experiments with lumen proteins such as plastocyanin (a small hydrophilic electron carrier) and the 33- and 23-kDa proteins of the photosynthetic oxygen-evolving complex have provided strong evidence for two separate steps in the transport process.[703,719,720,732,733] Moreover, these two distinct targeting events (viz., transport across the chloroplast envelope and subsequent translocation across the thylakoid membrane) are directed by a composite transit peptide that has two functionally independent domains (Figure 33). The bipartite structure is comprised of a N-terminal domain, which is structurally and functionally analogous to the transit peptides of stromal proteins, followed by a more hydrophobic domain (responsible for targeting across the thylakoid membrane).[701,702] Interestingly, the latter strongly resembles the signal sequences of secretory proteins in bacterial and eukaryotic cells.[744] It is noteworthy that precursor maturation also occurs in two sequential steps that require different proteases; one is located in the stroma and yields an intermediate protein form; the other is present as an integral membrane protein and generates the mature protein.[732,733,745–749] The thylakoid processing peptidase is capable of cleaving signal peptides of secretory proteins of both eukaryotic and bacterial origin (see Reference 701 and references therein). The site utilized by the stromal processing enzyme occurs within a central region of the transit peptide; the enzyme recognizes a motif that is common to the transit peptides of both stromal and thylakoid luminal proteins and is surrounded by a common secondary structure (predicted β-turn/β-sheet/α-helix structure). This structural similarity may be indicative of a selective pressure to maintain common sites on these precursors for recognition by the stromal processing peptidase.[750]

In vitro assays for the import of proteins by isolated thylakoids have been refined and optimized facilitating investigation of the specific requirements for targeting; in some cases import efficiencies approaching 100% have been achieved (e.g., for the 25- and 16-kDa proteins of the photosynthetic oxygen-evolving complex protein).[751] Use of the refined *in vitro* import system indicates that there may be distinct translocation systems involved in the transport of proteins across the thylakoid membrane, each recognizing specific features in the presequences of a subset of luminal proteins (see later discussion).

Cytochrome f is an example of a chloroplast-encoded protein that is targeted to the thylakoid membrane system; its transport is presumed to be functionally equivalent to that of imported stromal intermediates *en route* to the thylakoids. Following its synthesis in the stroma on thylakoid-bound ribosomes, the precursor inserts partially into the thylakoid membrane by a C-terminal stop-transfer domain; a transient N-terminal sequence is then removed following membrane insertion.[752]

C. Protein Targeting to the Envelope Membranes and Intermembrane Space

In comparison with the targeting of proteins to the stroma and thylakoids, much less is known about the targeting of chloroplast envelope proteins, in part because of their low quantities compared with other chloroplastic proteins. None of the outer envelope membrane proteins studied so far (e.g., 6.7- and 14-kDa proteins from spinach and pea, respectively) has a cleavable transit peptide.[753,754] Thus, proteins destined for the outer membrane of the chloroplastic envelope are likely to follow an import pathway distinct from that followed by proteins destined for other chloroplastic compartments, and the necessary information is most likely located within the mature protein (Figure 33), although its precise nature remains to be determined. The targeting of these proteins is unique in other ways; for example, it does not require ATP or a protease-sensitive receptor.[704]

By contrast, import studies with different inner membrane proteins (e.g., a 37-kDa protein and a phosphate translocator, both from spinach) indicate the presence of cleavable transit peptides.[755,756] Studies on the maize *bt-1*-encoded protein indicate that the transit peptides of inner membrane proteins function primarily as stromal targeting sequences; the specific information for subsequent targeting to the inner envelope is contained in the mature region of the protein.[757]

The components of the intermembrane space remain elusive. No proteins of this compartment have been identified, due in part to the inability to biochemically distinguish the intermembrane space from the stroma using current techniques.[704]

D. Energy Requirements

Energy is required for both protein translocation across the chloroplast envelope and for efficient binding of precursors to the outer envelope membrane.[758,759] Hydrolyzable nucleoside triphosphate (NTP) in the intermembrane space likely drives the binding of precursors to the outer-envelope receptor.[759,760] The site of NTP hydrolysis, the specificity of the requirement for NTP and the apparent K_m of the reaction distinguish the ATPase that mediates the binding reaction from what drives the translocation of bound precursor across the envelope membranes.[704,758–761] Hydrolyzable ATP (inside the chloroplast stroma) is the energy source utilized for translocation;[758] required concentrations for translocation are approximately five- to tenfold higher than those necessary for binding. Neither the electrical nor chemical components of a PMF are involved in translocation across the envelope membranes in chloroplasts.[758]

Efficient integration of LHCP into the thylakoid membrane requires energy;[734,762,763] as mentioned earlier, this requirement can be met by ATP alone, but the presence of a PMF across the thylakoid membrane in addition to ATP renders the process more efficient. Energy is also required for translocation across the thylakoid membrane into the lumen;[764] however, different precursors utilize different forms of energy for the translocation reaction[765,766] (reviewed in References 704 and 706) (Figure 34). This has led to the suggestion that thylakoid transfer signals direct translocation across the thylakoid membrane by two distinct mechanisms mediated by two distinct translocases in the thylakoid membrane.[706,766] As shown in Figure 34, precursors are imported into the stroma, probably by a common mechanism that requires ATP. From the stroma, translocation across the thylakoid membrane may occur via distinct pathways; translocation of plastocyanin and the 33-kDa subunit of the oxygen-evolving complex requires at least one stromal factor and ATP, but can take place in the absence of a PMF (the ΔpH component).[742,766] In contrast, translocation of the 23- and 16-kDa subunits of the oxygen-evolving complex requires only a ΔpH across the thylakoid membrane, with no requirement for ATP.[706,766] A third pathway is followed by the integral thylakoid protein, CF_0II (subunit II of the ATP synthase complex), whose insertion into the thylakoid membrane does not require stromal factors, ATP, or the ΔpH.[767] A spontaneous integration mechanism has been invoked for this protein (reviewed in Reference 706). Studies to deter-

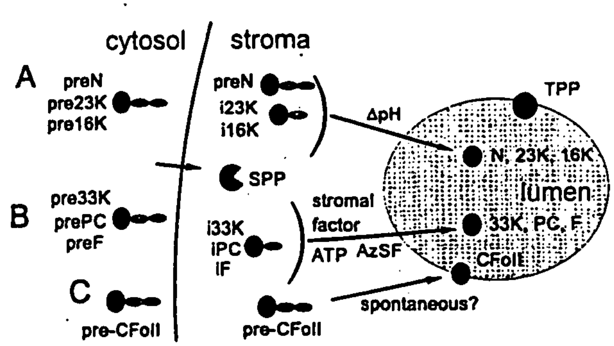


FIGURE 34. Multiple mechanisms of thylakoidal protein translocation. The six luminal proteins shown, and the integral membrane protein CFoII are synthesized with bipartite presequences and imported into the chloroplast, probably by a common mechanism. In the stroma, all of the precursors except pre-PSI-N and pre-CFoII are cleaved to intermediate forms by SPP. Further targeting into the thylakoids involves the operation of three distinct pathways. Translocation of PSI-N, 23K, and 16K across the thylakoid membrane appears to require only the thylakoidal ΔpH (pathway A), whereas translocation of 33K and PC is dependent on the presence of a stromal factor, ATP, and an azide-sensitive factor (AzSF); it is possible that the AzSF corresponds to the stromal translocation factor. The role of stromal factors and ATP in PSI-F translocation is not yet known, but this protein has been grouped with 33K and PC (pathway B) because translocation is sensitive to azide, does not require the thylakoidal ΔpH, and does not compete with that of 23K. A third pathway (C) is followed by the integral protein, CFoII, whose insertion into the thylakoid membrane does not require stromal factors, ATP or the ΔpH. After translocation into or across the thylakoid membrane, all of the proteins are apparently cleaved to the mature size by a common thylakoidal processing peptidase, TPP. (From Robinson, C. and Klösgen, R., *Plant Mol. Biol.*, 26, 15, 1994. With permission from Kluwer Academic Publishers.)

mine whether the different energy and stromal factor requirements reflect different translocation machineries in the thylakoid or whether a common translocation apparatus can operate in different energy-coupled modes[704] are awaited with interest.

E. Precursor Binding and the Identification of Receptors and Accessory Proteins

The first step in the transport of precursor proteins into chloroplasts is a specific interaction between the precursor and outer envelope membrane; the components that mediate this binding, as well as subsequent translocation, are likely to involve both membrane lipids and intrinsic proteins.[702] There are several lines of indirect evidence for receptor-mediated binding. For example, studies showing diminished binding of precursors following protease (thermolysin) treatment, which specifically destroys outer membrane proteins, indicate that a protein component(s) of the chloroplast outer envelope is necessary for high-affinity binding of precursors.[768,769] Moreover, precursor binding is ligand-specific; chloroplastic proteins lacking transit peptides do not bind to chloroplasts. Likewise, binding is membrane-specific; chloroplastic precursors generally do not bind to nonchloroplast membranes (e.g., plasma membranes and erythrocyte membranes) (see Reference 702 and references therein). Several other characteristics of precursor binding are indicative of receptor involvement (e.g., binding sites are limited and saturable; binding is rapid, specific, and requires energy in the form of ATP).[759,769,770] Saturation experiments led to an estimate of 1500 to 3500 binding sites (receptors) per chloroplast (reviewed in Reference 704).

At present, very little is known about the nature and composition of the translocation machinery. Translocation is thought to occur at regions where the inner and outer envelope membranes are closely appressed in so-called contact sites.[704,711] These sites may be dynamic structures, possibly formed as a consequence of protein translocation (reviewed in Reference 704). Several proteins that may be components of the signal recognition and transport apparatus have been identified. A heterobifunctional photoactivatable cross-linking reagent was used to implicate a 66-kDa chloroplast surface protein.[772] Experiments utilizing anti-idiotypic antibodies (in this case, antibodies capable of recognizing other antibodies specifically directed against a synthetic analog of the SSU-Rubisco transit peptide) have identified a putative (30-kDa) import receptor, apparently located at contact sites between the outer and inner envelope membranes.[773,774] However, recent work suggests that this protein is identical to the phosphate-3-phosphoglycerate-phosphate translocator,[775,776] suggesting a more indirect role in targeting.

ATP-dependent phosphorylation of a 51-kDa envelope protein is correlated with precursor translocation into chloroplasts.[777] However, whether it plays an integral role in the import process per se is unknown. Another approach for identifying putative transport components is to bind precursors to isolated outer membranes and then solubilize the membranes with detergent to identify a complex that remains associated with the bound precursor. Ten outer envelope proteins cosediment in a sucrose gradient with bound precursor SSU-Rubisco, including an 86-kDa protein, a Hsc70 protein, and a 34-kDa protein;[705,778] the complex isolated from the outer membrane can interact with precursor proteins.[779] More recently, a label transfer cross-linking reagent has been used to identify putative components of the transport apparatus;[780] the cross-linking was performed using precursor to the SSU-Rubisco that was blocked at an early stage of the transport process. In this approach, the cross-linker is cleavable, leaving the putative transport component modified with the radiolabeled portion of the cross-linker. Two envelope proteins were identified — an 86-kDa protein and a 75-kDa protein — both components of the outer membrane. Low levels of ATP stimulate labeling of the 75-kDa protein, whereas labeling of the 86-kDa protein is ATP-independent. The authors propose that precursors initially bind in an ATP-independent fashion to the 86-kDa protein; the energy-requiring step is associated with the 75-kDa protein and assembly of a translocation contact site between the inner and outer membrane of the chloroplastic envelope.[780] Protein-lipid interactions may also be involved in the initial binding of precursors to the chloroplast outer membrane; for example, they may mediate both the initial membrane insertion as well as facilitate transit peptide interaction with specific receptor proteins at the membrane surface.[702]

The role of molecular chaperones in the translocation process has been investigated; in carrying out their role, these proteins may reside on both sides of the membrane (reviewed in Reference 705). Hsp70 homologs have been identified within chloroplasts,[781,782] but their specific functions are not known. In higher plants, they are localized to the envelope[783] and stroma.[784,785] As mentioned earlier, proteins of the Hsp70 family are believed to prevent precursors from assuming a transport-incompetent conformation in the cytosol. However, a requirement for soluble factors in chloroplast protein import may be precursor specific, and little is known about the protein conformational changes that take place during precursor binding and import into chloroplasts. Some precursors of soluble stromal proteins are fully competent for import in the absence of cytosol (reviewed in Reference 704); other proteins such as LHCP may require cytosolic molecular chaperones to maintain an import-competent state.[786] Dihydrofolate reductase (DHFR) fusion proteins have been useful for studying conformational changes during protein import.[787,788] Folate analogs such as methotrexate bind tightly to DHFR and stabilize its mature folded conformation. This compound strongly inhibits the import of DHFR into yeast mitochondria,[789] presumably because the precursor must unfold to cross the membrane and is prevented from doing so when complexed

with methotrexate. In contrast, DHFR fusion proteins unfold despite the presence of methotrexate on binding to the chloroplast envelope. These results tend to support the existence of a strong protein unfolding activity associated with the chloroplast envelopes.[787,788] The role of chaperones located in the stroma in chloroplast protein import still needs to be fully addressed. The Reiske-Fe protein (a thylakoid membrane protein) becomes associated, in an ATP-dissociable form, with an Hsp70 homolog in the chloroplast stroma;[790] formation of a complex between the precursor and Hsp60 precedes the Hsp70 association. Recently a chloroplast homolog of the SRP (54-kDa subunit) has been identified in *Arabidopsis* that may play a role in intraorganellar targeting.[791] As noted earlier, several parallels exist between the targeting of chloroplast-encoded proteins to the thylakoids and the targeting of secretory proteins to the ER.

F. Assembly of Oligomeric Protein Complexes: Role of Chaperonins

As mentioned earlier, the synthesis and assembly of Rubisco involves the interaction of two genetic systems viz., nuclear and chloroplastic. The LSUs of Rubisco are synthesized within the chloroplast; the nuclear-encoded small subunits are imported after synthesis in precursor form on cytosolic polyribosomes.[707] The assembly of the holoenzyme (which in its mature assembled form consists of eight subunits of each type) occurs within the stromal compartment, and appears to require the presence of another protein, the Rubisco subunit binding protein, more recently termed the chloroplast chaperonin (cpn) (reviewed in References 792 through 794). This protein binds noncovalently to both newly synthesized LSUs and to imported small subunits. Convincing evidence for its role in assembly processes is available; for example, antiserum to the binding protein inhibits the transfer of newly synthesized Rubisco LSUs to the holoenzyme in stromal extracts (see Reference 792 and references therein). A representative of the chaperonin 60β gene family of *Arabidopsis* is both developmentally regulated and repressed by wounding when expressed in transgenic tobacco.[795] Interestingly, cpn is related to the GroEL protein of *Escherichia coli*, a heat-shock protein that is essential for cell growth[796] and bacteriophage assembly.[797] In addition, GroEL is required for assembly of prokaryotic Rubisco synthesized in *E. coli*[798] and it may also be involved in protein secretion.[799,800] Growth and development of transgenic tobacco plants transformed with the *E. coli groEL* gene are indistinguishable from control plants.[801] A second class of chaperonins (the 10-kDa class, Cpn10, equivalent to GroES in bacteria) has been identified recently in *Arabidopsis*.[802] Similar to the functions of the ER-resident molecular chaperones (discussed earlier), the cpns are proposed to assist other polypeptides in maintaining or assuming a conformation required for their correct assembly, possibly via their ATPase activities.[792,803] This role is not limited to the assembly of Rubisco; rather, the chloroplast chaperonin appears to be involved in the assisted assembly and/or folding of a wide range of proteins in chloroplasts.[804,805] Similar chaperonin-type proteins (e.g., Hsp60) have been implicated in oligomeric protein assembly in mitochondria (see later discussion). Thus, assisted assembly of oligomeric protein structures is emerging as a general cellular phenomenon and is an important aspect of regulation of gene expression at the posttranslational level.[792]

X. PROTEIN TARGETING INTO AND WITHIN MITOCHONDRIA

Mitochondria fulfill a variety of essential metabolic functions; in particular, as a major site for oxidative phosphorylation, they are often referred to as the "powerhouse" of the cell. The mitochondrial subcompartments each have a characteristic set of polypeptides and include the outer membrane, the intermembrane space, the inner membrane, and the matrix (Figure 32). Like the chloroplast, the great majority of mitochondrial proteins (i.e., greater than 90%) are contributed by the nucleocytosolic system.[806] Precursor synthesis on free polyribosomes of the cytosol is followed by posttranslational import into the organelle.[807,808] Much of our current understanding of the events required for organelle assembly and

mechanisms of mitochondrial protein transport has derived from *in vivo* and *in vitro* studies in animals, yeast, and *Neurospora* (reviewed in References 809 through 813). Much less information is available on mitochondrial protein transport in plants; however, some of the general principles derived from studies of other eukaryotic systems may also prove to be applicable to plants.

A. Protein Targeting to the Matrix

Proteins that are destined for the mitochondrial matrix (as well as most of those which will ultimately reside in the inner membrane or intermembrane space) are synthesized as precursors containing targeting sequences (termed "presequences") of 10 to 70 amino acids. In most cases (but not all),[814,815] these cleavable presequences are located at the N-terminus of the precursor. Much like the transit peptides of chloroplast precursors, sequence homology among mitochondrial presequences is lacking; common features include a high proportion of positively charged and hydroxylated residues and a general paucity of acidic amino acids.[816,817] Moreover, they also possess a high degree of autonomy, being both necessary and sufficient to direct nonmitochondrial passenger proteins into the mitochondrial matrix.[818,819] Recently, this has been shown for plant mitochondrial presequences in transgenic tobacco plants. The presequence of the mitochondrial β-subunit of F_1ATPase ($F_1\beta$) (from tobacco) is capable of directing the bacterial protein CAT into mitochondria, with high organellar specificity.[820] Likewise, the N-terminal 60 residues of the $F_1\beta$ presequence are sufficient for transport of cytosolic glutamine synthase (of *Phaseolus vulgaris*) to the mitochondria in transgenic tobacco.[821] When GUS is linked to the presequence of the δ-subunit of F_1ATPase (from sweet potato), the chimeric protein is efficiently transported into mitochondria only if the N-terminal regions of the mature protein are also present.[822] The nuclear-encoded precursor of mitochondrial superoxide dismutase (SOD isozyme 3, a manganese-containing homotetrameric enzyme) is translocated into isolated maize mitochondria. Deletions in the presequence of maize SOD-3 (generated *in vitro*) show relative import efficiency to be highly correlated with deletion size.[823]

A yeast mitochondrial presequence is capable of targeting a foreign protein into plant mitochondria *in vivo*.[824] The fusion protein consisted of the presequence of yeast mitochondrial tryptophanyl t-RNA-synthetase linked to GUS; specific targeting and efficient import into the mitochondria of transgenic tobacco cells occurred, with no substantial misrouting. Proteolytic processing of the precursor was equivalent in the two eukaryotic systems, with respect to both precision and fidelity; thus, the processing enzyme in plant mitochondria appears to recognize the same cleavage site within the presequence as the matrix protease from yeast.[824]

It appears then, that the targeting of chimeric proteins is organelle-specific; dual targeting into mitochondria and chloroplasts is not a general phenomenon. Exceptions do exist, however. For example, a yeast mitochondrial presequence (linked to a bacterial passenger protein) is recognized and can interact functionally with the protein translocation systems of both chloroplasts and mitochondria in transgenic tobacco.[825]

It is generally assumed that the essential feature of the presequence in the binding/translocation process is encrypted in a specific secondary or tertiary conformation (rather than in specific amino acid residues). Many targeting signals of mitochondrial precursors have the potential to form amphiphilic α-helices and β-sheets (i.e., secondary structural arrangements with the polar and nonpolar residues exposed to opposite faces).[729,730,826] Such structures are said to have a hydrophobic moment,[827] and react spontaneously with the surfaces of biological membranes.[828] This amphiphilicity is proposed to be important for the initial membrane insertion and/or the interaction with specific receptor proteins at the mitochondrial surface.[811]

Most protein translocation into mitochondria occurs at sites where the outer and inner membranes are in close proximity.[829] It proceeds through a hydrophilic (proteinaceous) membrane environment; protein-lipid interactions may also occur.[830] The translocation complex seems to consist of two distinct translocation channels in the outer and inner membranes (see later discus-

sion).[813,831] Once translocated across the mitochondrial membranes, the N-terminal presequences of precursors are cleaved by a highly specific metal-dependent processing enzyme in the matrix.[832] This proteolysis is not coupled to membrane translocation, but is an essential process (e.g., in yeast),[833] probably being required for proper assembly of imported proteins.[811] Proteolytic processing is thought to be a cooperative effort requiring two structurally related components that may recognize different structural elements of presequences, and hence, contribute to the high specificity of cleavage. One component may bind to presequences of incoming proteins (in a cotranslational manner), thus exposing the cleavage site towards the other component.[811,834] In the mitochondria of lower eukaryotes and mammals, the general processing peptidase consists of two subunits, α-MPP and β-MPP, both located in the matrix[808,834,836–838] (reviewed in Reference 835). In *Neurospora*, the β-MPP is primarily membrane-associated, although the active part of this subunit is located in the matrix.[839] In potato and spinach, the processing enzyme is an integral part of the bc_1 complex;[840,841] processing activity of spinach leaf mitochondria is membrane bound.[842] There may be three components of the mitochondrial processing enzyme of potato.[840] Interestingly, there are tissue-specific differences in the activity and solubility of the general processing enzyme in mitochondria of spinach leaves vs. roots; two populations of the processing peptidases may be present in root mitochondria.[843]

B. Protein Targeting to the Inner Membrane and Intermembrane Space

Proteins residing in the matrix reach their target compartment by translocation across the two membranes at contact sites; additional routing is required for correct localization of proteins of the inner membrane or intermembrane space. Soluble proteins of the intermembrane space (e.g., cytochrome b_2 of yeast mitochondria) as well as cytochrome c_1 of the bc_1 complex (which is largely exposed to the intermembrane space) are synthesized as cytosolic precursors with long complex presequences[844–846] having a bipartite structure. Their N-terminal parts exhibit the typical features of presequences of matrix-targeted proteins (and are functionally equivalent); the remaining C-terminal parts contain numerous hydrophobic residues preceded by one to four basic residues. This latter motif is thought to direct "export" of intermembrane proteins from the matrix back across the inner membrane.[847,848] However, another view is that it functions as a stop-transfer signal. According to the stop-transfer hypothesis, the N-terminal part of the presequence is imported into the matrix, but the intermembrane space targeting domain arrests further translocation through the inner membrane (reviewed in Reference 813). The composite presequences undergo cleavage in two steps executed by different processing peptidases: the matrix α-MPP/β-MPP and a membrane-associated peptidase at the outer surface of the inner membrane (IMP I, inner membrane peptidase).[835,847,849–851] Cytochrome c_1 from potato also contains a transient presequence with a bipartite structure comparable to that described for the fungal proteins;[840] the targeting mechanism of the plant protein has not yet been determined.

The Rieske Fe-S protein is a peripheral component of the bc_1 complex located at the outer surface of the mitochondrial inner membrane; in yeast and *Neurospora*, the protein reaches its functional location on an import route via the matrix.[811,852,853] However, the presequence lacks a hydrophobic segment and contains information for targeting to the matrix only; information for its subsequent relocation to its final destination may be encoded in the mature protein.[854,855] Proteolytic processing is in the matrix only, and occurs via a two-step mechanism. The Reiske Fe-S protein of potato shares about 50 and 43% sequence identity with the same proteins from fungi and mammals, respectively.[856] The plant protein is made as a larger precursor of 30 kDa; its presequence of 53 amino acids has molecular features different from those found in presequences of fungal iron-sulfur proteins. Further, unlike in yeast and *Neurospora*, the presequence of the plant protein is removed by a single processing step.[856]

It is noteworthy that there are a few proteins of the mitochondria whose assembly and transport pathways are unique and do not conform to conventional intramitochondrial targeting routes (i.e., rerouting following transport into the matrix).[811] The ADP- and ATP-carrier (AAC) of the inner membrane (and possibly some other structurally-related proteins) are examples. The cytosolic precursor of AAC is made without a cleavable presequence;[814] targeting information resides in three internal segments.[857] Following its entry into the outer membrane (and subsequent transport into contact sites), there is lateral diffusion into the inner membrane. The signals that prevent translocation into the matrix and trigger integration into the inner membrane are unknown.[811] Cytochrome c also does not follow a reexport pathway; it reaches the intermembrane space by crossing the outer membrane only.[858,859]

C. Other Components Involved in Protein Import and Assembly

Several of the components of the mitochondrial import machinery have been identified recently, mostly in *Neurospora crassa* and in the yeast *S. cerevisiae*. These include (1) cytosolic cofactors, particularly chaperone components; (2) import receptors and accessory components of the outer membrane; (3) components of the translocation machinery of the inner membrane; and (4) the matrix heat-shock proteins Hsp70, Hsp60, and their partners (reviewed in Reference 809) (Figure 35).

Precursor proteins must be maintained in a loosely folded conformation after synthesis, for their subsequent efficient translocation across mitochondrial membranes.[860–862] Constitutively expressed heat-shock proteins of the Hsp70 family have been implicated in the stabilization of precursors in a translocation-competent state; by binding to precursors cotranslationally, they may help to keep targeting signals exposed.[107,124,863–867] A NEM-sensitive factor associated with the cytosolic surface of mitochondria also appears to play a cooperative, facilitative role; along with ATP, it is required for the release of proteins from Hsc70.[868,869] As mentioned earlier, the translocation complex consists of two distinct translocation channels in the outer and inner membranes (Figure 35).[809,870] The first specific step of the import pathway is the binding of precursors to the surface of the mitochondrial outer membrane.[871,872] In *Neurospora*, two outer membrane receptor/binding proteins (MOM19 and MOM72 — 19 and 72 kDa, respectively) have been identified.[873] Inhibition of precursor binding and import by antibodies recognizing the cytosolic domains of the proteins is indicative of their role as receptors that participate in the specific recognition of precursors; also significant is the observation that these proteins are enriched in the outer membrane at contact sites between inner and outer membranes, where most translocation takes place.[873] The subsequent entry of precursors into the translocation apparatus is thought to be facilitated by a common (integral) component in the outer membrane, the general insertion pore (GIP) (Figure 35) (see Reference 870 and references therein). Four subunits (MOM38, MOM30, MOM8 and MOM7) may constitute the GIP. These are in close proximity to precursors arrested in the GIP and presumably contribute to formation of the pore. Yet another component may be needed for the transfer of preproteins from the surface-exposed receptors MOM19 and MOM72 to the membrane-embedded components of the GIP. This appears to be the role of MOM22 — a central component of the receptor complex whose negatively charged domain (exposed to the cytosol) may allow it to interact with and transfer the positively charged presequences on preproteins (Figure 36).[870] Proteins of the matrix, inner membrane, and intermembrane space may be routed from the GIP into contact sites for further translocation; precursors of outer membrane proteins (e.g., porin) may also insert into their target membrane directly via the GIP.[811]

After translocation of precursors through the mitochondrial innermembrane import machinery (MIM), the presequence is proteolytically removed in the matrix, and proteins destined to reside in the matrix are folded into their functional forms (Figure 36). Current models envision that the import systems of the outer and inner membranes can be transiently linked by translocating polypeptide chains,[831] thereby forming translocation contact sites.[829] Thus, the two transport machineries are not permanently connected. Further, translocation of preproteins across the outer membrane

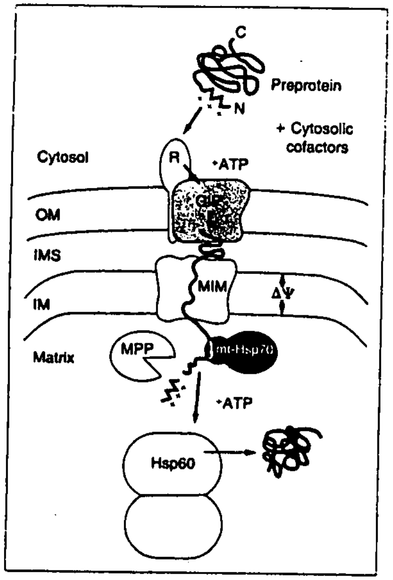


FIGURE 35. Basic scheme of mitochondrial protein import. The transport pathway of a preprotein into the mitochondrial matrix is shown. The preprotein typically carries a positively charged presequence at its amino terminus (N). Cytosolic cofactors prevent misfolding or aggregation; some of them function in an ATP-dependent manner. The preprotein binds to a receptor (R) and is translocated across the outer membrane (OM) through the general insertion pore (GIP). Translocation across the inner membrane (IM) is mediated by the mitochondrial inner membrane import machinery (MIM), and requires a membrane potential (ΔΨ). In the matrix, the heat-shock protein Hsp70 binds to the preprotein; the mitochondrial processing peptidase MPP cleaves off the presequence, and the protein is folded, aided by Hsp60. Both heat-shock proteins require ATP. The outer- and the inner-membrane machineries are not permanently connected by a sealed channel, but a preprotein in transit across the membranes contacts the intermembrane space (IMS). However, the outer- and inner-membrane machineries can be transiently connected by a preprotein spanning both membranes. C, carboxyl terminus. (From Pfanner, N., Craig, E. A., and Meijer, M., *Trends Biochem. Sci.*, 19, 368, 1994. With permission from Elsevier Publishing, Cambridge.)

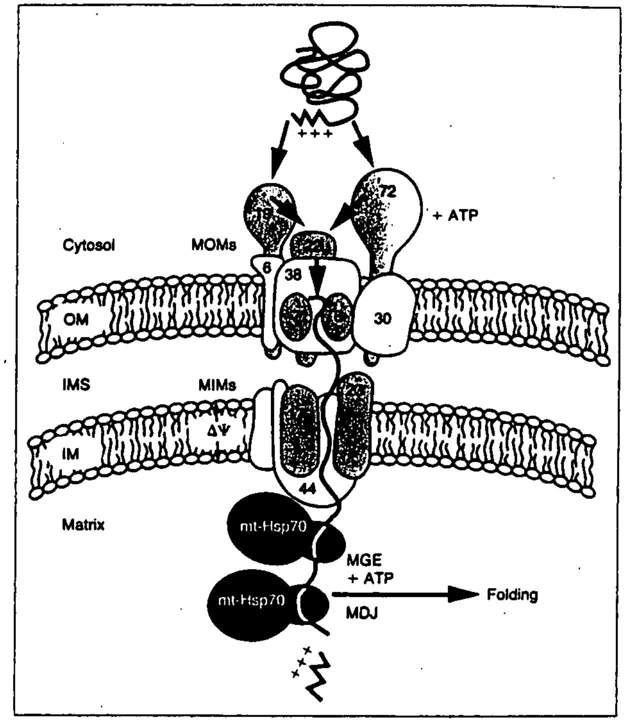


FIGURE 36. Hypothetical model of the protein-import machineries in the mitochondrial membranes. The mitochondrial outer membrane (MOM) machinery consists of at least eight different proteins: the import receptors MOM19 and MOM72 (MOM proteins of 19 and 72 kDa, respectively); the transfer component MOM22; the general insertion pore (GIP) with MOM38, MOM30, MOM8, and MOM7; and ISP6, which was shown to interact genetically with MOM38/ISP42. The integral membrane proteins MIM17 and MIM23 and the preprotein-binding protein MIM44 are components of the mitochondrial inner-membrane import machinery (MIM). The mitochondrial Hsp70 (mt-Hsp70) also binds the polypeptide chain in transit. Mt-Hsp70 is assumed to interact with MGE and MDJ, the mitochondrial homologs of the prokaryotic heat-shock proteins GrpE and DnaJ. Interaction of the positively charged presequence with the MIM machinery requires a membrane potential, (ΔΨ). IM, inner membrane; IMS, intermembrane space; OM, outer membrane. (From Pfanner, N., Craig, E. A., and Meijer, M., *Trends Biochem. Sci.*, 19, 368, 1994. [Also see references therein.] With permission from Elsevier Publishing, Cambridge.)

is independent of the membrane potential ($\Delta\Psi$); in contrast there is a strict requirement for a $\Delta\Psi$ for translocation across the inner membrane.[809,874] Three MIM proteins are essential components of the import machinery in yeast and *Neurospora* (reviewed in Reference 809) (Figure 36). Two of these components (MIM17 and MIM23/MAS6) are thought to form part of a channel. The third essential component, MIM44, in cooperation with the heat-shock protein Hsp70, binds the preproteins in transit. The latter (i.e., matrix-Hsp70) plays a dual role in the translocation process.[875–877] First, by binding directly to the preprotein, it facilitates the stepwise unfolding of the remainder of the polypeptide chain on the outside of the outer membrane by trapping preprotein segments emerging on the matrix side (unfoldase function). Second, it is essential for the complete import of preproteins into the matrix, independent of their folding state (translocase function). In cooperation with the mitochondrial homologs of bacterial DnaJ and GrpE, Hsp70 forms an initial part of the machinery for folding imported proteins by transferring them to Hsp60 (see below)[809] (Figure 36). The MIM machinery is likely not a static channel; rather, the MIM proteins and their partners, such as the matrix Hsp70, may assemble and disassemble in a dynamic manner.[809] Both an electrical membrane potential and ATP are needed to drive protein translocation through the MIM import machinery.[809] The membrane potential, $\Delta\Psi$, triggers the translocation of the initial part of the preprotein, the N-terminal presequence, across the inner membrane.[878] ATP in the matrix is needed to complete the translocation of the presequence and to allow translocation of the mature portion into the matrix rendering the import process irreversible.[271,864,875,877] Although part of this ATP requirement is due to the ATP-dependent action of the matrix-Hsp70, further components of the import machinery are likely to require ATP to function.[809] Hsp60 (a large protein complex located in the matrix) plays an integral role in protein folding and assembly *in vivo*.[792,811,879–881] This protein may have several functions. It mediates the ATP-dependent folding of imported matrix proteins; folding may be prevented until the complete polypeptide is available.[880] A role in the maintenance of a translocation-competent state is also suggested, particularly for imported proteins that must undergo further intraorganellar targeting (e.g., insertion into, or translocation across, the inner membrane).[879] Finally, Hsp60 of the mitochondrial matrix is highly homologous to the GroEL protein of *E. coli* and to the Rubisco-binding protein of chloroplasts, and is also a member of the subgroup of molecular chaperones termed "chaperonins" (see earlier discussion). Thus, related to its role in mediating protein folding (e.g., recognizing structural motifs in unfolded or loosely folded polypeptide chains and repairing misfolded proteins), Hsp60 may also play an essential role in the assembly of large oligomeric proteins in the matrix.[792]

XI. PROSPECTS FOR ENGINEERING CHLOROPLAST AND MITOCHONDRIAL PROTEINS

A recent focus of genetic engineering in plants has been toward the ultimate development of systems for influencing the efficiency of plant growth and carbon fixation by photosynthesis, as well as the engineering of herbicide tolerance. This has involved the use of many novel approaches, including the manipulation of genes encoding important proteins (e.g., key enzymes of metabolic pathways) coupled with the exploitation of targeting mechanisms of proteins within both cells and organelles.

A. Herbicide Tolerance

Genetic engineering of herbicide tolerance into crop species is of significant interest to agricultural biotechnology. Particular interest in engineering tolerance to the herbicide glyphosate (*N*-[phosphono-methyl] glycine) stems from some of its desirable properties (e.g., its nontoxicity to animals and rapid degradation by soil microorganisms).[882] This herbicide functions by inhibiting the enzyme 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase, and thus prevents plant growth by blocking the pathway of aromatic amino acid synthesis.[883] Some degree of tolerance to this herbicide has been achieved by introducing into host plants a bacterial EPSP synthase gene (that encodes a resistant enzyme). To direct correct chloroplastic localization of this mutant enzyme form, the DNA encoding the plant EPSP synthase transit peptide was linked to the bacterial enzyme coding region. The *in vitro* product of this chimeric gene is rapidly imported into chloroplasts, where it accumulates as a stable, glyphosate-resistant enzyme.[722] Another strategy has been to introduce a plant EPSP synthase gene under the control of the cauliflower mosaic virus (CaMV) 35S promoter; in this case, the inhibitory effect of glyphosate is counteracted by overproduction of the plant enzyme targeted to the chloroplast by its own transit peptide.[884]

A further approach to the engineering of herbicide tolerance involves a recent technology developed for the conversion of chloroplast (and mitochondrial) genes into nuclear genes and for the subsequent retargeting of the cytosolically synthesized protein into its respective "home" organelle (reviewed in Reference 885). This effective relocation of genes from organelle to nucleus mimics the natural events proposed to have occurred during evolution (i.e., the relocation of the majority of genes from an endosymbiotic organism to the host cell nucleus). Reformatting organellar genes for this "allotopic" expression (i.e., expression in a foreign or alien environment) involves several manipulations, including:

1. Changing the open reading frame of the organellar gene (e.g., so that codon usage is compatible with the biases of the nucleo-cytosolic system).
2. Placing the open reading frame within a suitable nuclear transcription context (including appropriate promoter and terminator); a functional nuclear replicator sequence is provided to ensure the replication and maintenance of the restructured gene in the nucleus.
3. Addition of DNA encoding a N-terminal leader sequence to ensure correct targeting of the fusion protein into its home organelle, subsequent to its synthesis in the cytosol.

This strategy has been used in attempts to engineer resistance to the triazine herbicides (e.g., atrazine). The triazine group of herbicides compete with thylakoid plastocyanin for binding with the quinone-binding protein (Q_B of photosystem II) and, thus, interrupt photosynthetic electron flow.[886] In thylakoid membranes, the Q_B protein of atrazine-sensitive plants, including the weed *Amaranthus hybridus*, binds azidoatrazine; Q_B protein from an atrazine-resistant biotype does not bind the herbicide because of a single amino acid residue substitution.[887] The allotopic expression of the resistant form of Q_B protein from *Amaranthus* has been achieved in host tobacco cells.[888] The coding region of the donor (resistant) gene (*psbA*) was fused to the promoter and transit-peptide-encoding sequences of the SSU-Rubisco gene, and the chimeric gene was introduced into tobacco plants. The transgenic tobacco plants were tolerant to levels of atrazine that are toxic to nontransformed plants, although the atrazine-resistant phenotype was not sustained for long periods of plant growth.[888] Because expression of the endogenous ("sensitive") gene was not blocked in these plants, the variant Q_B protein was delivered to the organelles in which normal levels of the naturally encoded protein were produced. Thus, the transformants contained a mixed population of photosystem II complexes. Nonetheless, correct targeting of the variant Q_B form to the chloroplast could be demonstrated (by immunocytochemical means), and the recovery of atrazine-tolerant transgenic plants shows that the protein functions in photosynthesis.[888]

B. Increasing Photosynthetic Productivity

Rubisco catalyzes the first step in the processes of both photosynthesis and photorespiration; because it is the balance between these two processes that ultimately controls plant productivity, this enzyme is a major target for genetic engineering (e.g., mutagenesis), with the aim of altering this balance for agricultural purposes.[889] Toward this goal, expression of Rubisco in a prokaryotic host has been undertaken.[792] Until recently, little success has been achieved, in part due to a failure of Rubisco subunits to assemble properly in the bacterial host; hence, an enzymatically active form is not produced.[792,890] Assembly may not occur because the chaperonins of these bacterial host cells, which would normally medi-

ate protein assembly, are sufficiently different from those of plant cells, and hence, are nonfunctional in Rubisco assembly. Thus, a current aim is to express the cDNAs for the chloroplast chaperonin in the same *E. coli* cells that are expressing the genes for Rubisco from higher plants. In this way, Rubisco assembly should be rescued, permitting attempts to improve the properties of this agriculturally important enzyme.[792]

Allotopic expression of the LSU of Rubisco has been achieved in *Oenothera hookeri* (evening primrose) (see Reference 885 and references therein). The transit peptide and transcriptional control sequences from the pea SSU-Rubisco gene were utilized for nucleocytosolic expression; the host was a plastome mutant containing a sigma mutation[891] in the Rubisco LSU (*rbcL*) gene of chloroplast DNA. Hence, the mutant expresses only a truncated version of the LSU. Allotopic expression of full-length LSU was achieved, curing the sigma phenotype in transformants.

Because the technology of allotopic expression allows both controls over quantitative expression of the engineered gene and the subsequent correct delivery to the organelle of a protein of defined structure, it should have significant potential as a novel approach for basic investigation into the biogenesis, assembly, and function of a number of enzyme complexes in the chloroplast.[885] Its potential as a natural assay system for directly manipulating photosynthetic productivity (e.g., via changes to Rubisco subunits to diminish photorespiration and enhance CO_2 fixation) may also be exploited in the future.

C. Allotopic Expression of Mitochondrial Genes

This strategy may also prove to be useful for similar basic and applied studies of mitochondrial-encoded proteins. Successful allotopic expression of at least two mitochondrial genes has been achieved in yeast.[885] For example, expression of a yeast gene encoding subunit 8 of mitochondrial ATPase was demonstrated in mutant host cells unable to synthesize the endogenous mitochondrial protein; transit peptide sequences (derived from the mitochondrial *Neurospora* subunit 9 gene) effected the required mitochondrial import.[892] In a similar manner, another mitochondrial gene (for the intron-encoded maturase of cytochrome b) was expressed allotopically in yeast cells.[893]

Thus, there appears to be no intrinsic biological barrier to chloroplast or mitochondrial genes being encoded in the nucleus; the subsequent import of the synthesized protein into the target organelle is also possible, provided an appropriate presequence or transit peptide is utilized.[885] Other approaches to the manipulation and analysis of organellar genes could potentially make use of current techniques for direct delivery of organellar DNA into mitochondria and chloroplasts (e.g., via projectile techniques, polyethylene glycol [PEG] treatment, or possibly via DNA-protein conjugates).[894–898] Although these approaches may have a more limited potential than allotopic expression strategies,[885] future attempts to engineer organellar genes and their encoded protein products may exploit a combination of these avenues.

XII. SUMMARY

It is apparent that posttranslational controls are an important aspect of the regulation of gene expression in all eukaryotes. Studies on the mechanisms of protein targeting in plant cells have progressed markedly in the last 5 years; however, in certain areas, much more information is presently available in other eukaryotic systems. Although important exceptions exist, a striking feature of the mechanisms and cellular machinery of protein targeting is their universality — among animals, plants, and eukaryotic microorganisms — and even between prokaryotes and eukaryotes. Mechanisms of vesicle targeting and the role of GTPases will very shortly become a rapidly developing aspect of plant cell biology. Studies have also progressed toward characterizing signals involved in the transport and targeting of proteins to the nucleus, chloroplast, mitochondrion, peroxisome, and glyoxysome. The recent refinement of methods for subcellular localization of specific proteins (and their corresponding transcripts) will aid further progress in this area.[899] The specificity of protein-targeting processes is the very basis of maintaining structural and functional integrity of the cell, enabling the various subcellular compartments to carry out their diverse metabolic roles.

Protein stability may also be influenced by subcellular localization; various factors appear to contribute to a protein's half-life in the cell. One of these is the ability of the protein to undergo normal folding and assembly processes and to assume a functional three-dimensional structure; failure to achieve this results in the activation of the cell's disposal mechanisms to remove nonfunctional proteins or components. A consideration of these factors also has far-reaching implications for applied studies geared toward the genetic engineering of plants for agronomically useful traits or characteristics, in particular, where success will depend on achieving high levels of accumulation of a foreign protein in a heterologous host. The ultimate challenge here is to design or manipulate polypeptide sequences (e.g., toward such desirable characteristics as enhanced nutritive value or more productive metabolic characteristics), but maintain a functional and stable three-dimensional conformation; this is impeded by our limited knowledge of the rules that govern protein folding and oligomer assembly and how these processes relate to a protein's ultimate stability in the cell. Thus, in addition to a characterization of protein targeting signals, more information is required about the structural features of proteins that allow for their stable accumulation in a particular subcellular compartment.

ACKNOWLEDGMENTS

I am grateful to Gad Galili, David G. Robinson, Brian W. Tague, Kenzo Nakamura, Roger Beachy, Jean-Marc Neuhaus, Bud Ryan, Faith Belanger, Jon T. Lindstrom, Aldo Ceriotti, Gerhard Saalbach, John J. Harada, Laura J. Olsen, Leonard Beevers, Maarten J. Chrispeels, Robert P. Donaldson, Natasha Raikhel, and Alan B. Bennett for making available their most recent data. Many thanks are also offered to Diana Elm for her prompt and accurate typing of the manuscript.

This article is dedicated to the memory of Lyle and Helen Anderson, who lived to have fun and show kindness to others, and to their surviving sons Gary and Wayne Anderson.

REFERENCES

1. **Della-Cioppa, G., Kishore, G. M., Beachy, R. N., and Fraley, R. T.**, Protein trafficking in plant cells, *Plant Physiol.*, 84, 965, 1987.
2. **Battey, N. H. and Blackbourn, H. D.**, The control of exocytosis in plant cells, *New Phytol.*, 125, 307, 1993.
3. **Satiat-Jeunemaitre, B. and Hawes, C.**, Insights into the secretory pathway and vesicular transport in plant cells, *Biol. Cell*, 79, 7, 1993.
4. **Vitale, A., Ceriotti, A., and Denecke, J.**, The role of the endoplasmic reticulum in protein synthesis, modification and intracellular transport, *J. Exp. Bot.*, 44, 1417, 1993.
5. **Verner, K. and Shatz, G.**, Protein translocation across membranes, *Science*, 241, 1307, 1988.
6. **Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D.**, *Molecular Biology of the Cell*, 2nd ed., Garland Publishing, New York, 1989.
7. **Pelham, H. R. B. and Munro, S.**, Sorting of membrane proteins in the secretory pathway, *Cell*, 75, 603, 1993.
8. **Blobel, G.**, Intracellular protein topogenesis, *Proc. Natl. Acad. Sci. U.S.A.*, 77, 1496, 1980.
9. **Wessels, H. P. and Spiess, M.**, Insertion of a multispanning membrane protein occurs sequentially and requires only one signal sequence, *Cell*, 55, 61, 1988.
10. **Austen, B. M. and Westwood, O. M. R.**, *Protein Targeting and Secretion*, Oxford University Press, Oxford, 1991.
11. **von Heijne, G. and Gavel, Y.**, Topogenic signals in integral membrane proteins, *Eur. J. Biochem.*, 174, 671, 1988.
12. **Kermode, A. R.**, Regulation of plant gene expression at the post-translational level: applications to genetic engineering, in *Mechanisms of Plant Growth and Improved Productivity. Modern Approaches*, Basra, A. S., Ed., Marcel Dekker, New York, 1994, 317.
13. **Farrell, L. B. and Beachy, R. N.**, Manipulation of β-glucuronidase for use as a reporter in vacuolar targeting studies, *Plant Mol. Biol.*, 15, 821, 1990.
14. **Pugsley, A. P.**, *Protein Targeting*, Academic Press, San Diego, CA, 1989.
15. **Both, G.**, Protein transport, *Today's Life Science*, August, 12, 1990.
16. **Pfeffer, S. R. and Rothman, J. E.**, Biosynthetic protein transport and sorting by the endoplasmic reticulum and Golgi, *Annu. Rev. Biochem.*, 56, 829, 1987.

17. **Chrispeels, M. J.**, Sorting of proteins in the secretory system, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 42, 21, 1991.
18. **Wink, M.**, The plant vacuole: a multifunctional compartment, *J. Exp. Bot.*, 44 (Suppl.), 231, 1993.
19. **Cyr, D. R. and Bewley, J. D.**, Proteins in the roots of the perennial weeds chicory (*Cichorium intybus* L.) and dandelion (*Taraxacum officinale* Weber) are associated with overwintering, *Planta*, 182, 370, 1990.
20. **Baba, K., Ogawa, M., Nagano, A., Kuroda, H., and Sumiya, K.**, Developmental changes in the bark lectin of *Sophora japonica* L., *Planta*, 183, 462, 1991.
21. **Chrispeels, M. J. and Raikhel, N. V.**, Lectins, lectin genes and their role in plant defense, *Plant Cell*, 3, 1, 1991.
22. **Sonnewald, U., Studer, D., Rocha-Sosa, M., and Willmitzer, L.**, Immunocytochemical localization of patatin, the major glycoprotein in potato (*Solanum tuberosum* L.), *Planta*, 178, 176, 1989.
23. **Wetzel, S., Demmers, C., and Greenwood, J. S.**, Seasonally fluctuating bark proteins are a potential form of nitrogen storage in three temperate hardwoods, *Planta*, 178, 275, 1989.
24. **Staswick, P. E.**, Novel regulation of vegetative storage protein genes, *Plant Cell*, 2, 1, 1990.
25. **Bol, J. F., Linthorst, H. J. M., and Cornelissen, B. J. C.**, Plant pathogenesis-related proteins induced by virus infection, *Annu. Rev. Phytopathol.*, 28, 113, 1990.
26. **Van den Bulcke, M., Bauw, G., Castresana, C., Van Montagu, M., and Vandekerckhove, J.**, Characterization of vacuolar and extracellular β (1,3)-glucanases of tobacco: evidence for a strictly compartmentalized plant defense system, *Proc. Natl. Acad. Sci. U.S.A.*, 86, 2673, 1989.
27. **Knox, J. P.**, Molecular probes for the plant cell surface, *Protoplasma*, 167, 1, 1992.
28. **Cassab, G. I. and Varner, J. E.**, Cell wall proteins, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 39, 321, 1988.
29. **Jones, R. L. and Robinson, D. G.**, Protein secretion in plants, *New Phytol.*, 111, 567, 1989.
30. **Showalter, A. M.**, Structure and function of plant cell wall proteins, *Plant Cell*, 5, 9, 1993.
31. **Varner, J. E. and Lin, L.-S.**, Plant cell wall architecture, *Cell*, 56, 231, 1989.
32. **Roberts, K.**, *Curr. Opin. Cell Biol.*, 1, 1020, 1989.
33. **Ye, Z.-H. and Varner, J. E.**, Tissue-specific expression of cell wall proteins in developing soybean tissues, *Plant Cell*, 3, 23, 1991.
34. **Jones, R. L., Gilroy, S., and Hillmer, S.**, The role of calcium in the hormonal regulation of enzyme synthesis and secretion in barley aleurone, *J. Exp. Bot.*, 44 (Suppl.), 207, 1993.
35. **Kermode, A. R.**, Regulatory mechanisms involved in the transition from seed development to germination, *Crit. Rev. Plant Sci.*, 9, 155, 1990.
36. **Kermode, A. R.**, Regulatory mechanisms in the transition from seed development to germination: interactions between the embryo and the seed environment, in *Seed Development and Germination*, Galili, G. and Kigel, J., Eds., Marcel Dekker, New York, 1994, 273.
37. **Jacinto, T., Farmer, E. E., and Ryan, C. A.**, Purification of potato leaf plasma membrane protein pp34, a protein phosphorylated in response to oligogalacturonide signals for defense and development, *Plant Physiol.*, 103, 1393, 1993.
38. **Michelet, B., Lukaszewicz, M., Dupriez, V., and Boutry, M.**, A planta plasma membrane proton-ATPase gene is regulated by development and environment and shows signs of translational regulation, *Plant Cell*, 6, 1375, 1994.
39. **Maurel, C., Reizer, J., Schroeder, J. I., and Chrispeels, M. J.**, The vacuolar membrane protein γ-TIP creates water specific channels in *Xenopus* oocytes, *EMBO J.*, 12, 2241, 1993.
40. **Johnson, K. D., Herman, E. M., and Chrispeels, M. J.**, An abundant, highly conserved tonoplast protein in seeds, *Plant Physiol.*, 91, 1006, 1989.
41. **Bethke, P. C. and Jones, R. L.**, Ca^{2+}-Calmodulin modulates ion channel activity in storage protein vacuoles of barley aleurone cells, *Plant Cell*, 6, 277, 1994.
42. **DeLisle, A. J. and Crouch, M. L.**, Seed storage protein transcription and mRNA levels in *Brassica napus* during development and in response to abscisic acid, *Plant Physiol.*, 91, 617, 1989.
43. **Wilen, R. W., Mandel, R. M., Pharis, R. P., Holbrook, L. A., and Moloney, M. M.**, Effects of abscisic acid and high osmoticum on storage protein expression in microspore embryos of *Brassica napus*, *Plant Physiol.*, 94, 875, 1990.
44. **Wilen, R. W., van Rooijen, J. H., Pearce, D. W., Pharis, R. P., Holbrook, L. A., and Moloney, M. M.**, Effects of jasmonic acid on embryo-specific processes in *Brassica* and *Linum* oilseeds, *Plant Physiol.*, 95, 399, 1991.
45. **Jiang, L., Downing, W., Baszczynski, C., and Kermode, A. R.**, The 5′ flanking regions of vicilin and napin storage protein genes are down-regulated by desiccation in transgenic tobacco, *Plant Physiol.*, 107, 1439, 1995.
46. **Reinbothe, S., Mollenhauer, B., and Reinbothe, C.**, JIPs and RIPs: the regulation of plant gene expression by jasmonates in response to environmental cues and pathogens, *Plant Cell*, 6, 1197, 1994.
47. **Mason, H. S. and Mullet, J. E.**, Expression of two soybean vegetative storage protein genes during development and in response to water deficit, wounding, and jasmonic acid, *Plant Cell*, 2, 569, 1990.
48. **Mason, H. S., DeWald, D. B., Creelman, R. A., and Mullet, J. E.**, Coregulation of soybean vegetative storage protein gene expression by methyl jasmonate and soluble sugars, *Plant Physiol.*, 98, 859, 1992.
49. **Jones, R. L. and Jacobsen, J. V.**, Regulation of synthesis and transport of secreted proteins in cereal aleurone, *Intl. Rev. Cytol.*, 126, 49, 1991.
50. **Somerville, C. R.**, How is the amount of a membrane regulated?, in *Molecular Approaches to Compartmentation and Metabolic Regulation*, Huang, A. H. C. and Taiz, L., Eds., The American Society of Plant Physiologists, Rockville, MD, 1991, 151.

51. **Hepler, P. K., Palevitz, B. A., Lancelle, S. A., McCauley, M. M., and Lichtscheidl, I.**, Cortical endoplasmic reticulum in plants, *J. Cell Sci.*, 96, 355, 1990.
52. **Grabski, S., de Feijter, A. W., and Schindler, M.**, Endoplasmic reticulum forms a dynamic continuum for lipid diffusion between contiguous soybean root cells, *Plant Cell*, 5, 25, 1993.
53. **Quader, H., Hofmann, A., and Schnepf, E.**, Reorganization of the endoplasmic reticulum in epidermal cells of onion bulb scales after cold stress: involvement of cytoskeletal elements, *Planta*, 177, 273, 1989.
54. **Blobel, G. and Dobberstein, D.**, Transfer of proteins across membranes. II. Reconstitution of functional rough microsomes from heterologous components, *J. Cell Biol.*, 67, 835, 1975.
55. **Cutler, D. F.**, The role of transport signals and retention signals in constitutive export from animal cells, *J. Cell Sci.*, 91, 1, 1988.
56. **Burgess, T. L. and Kelly, R. B.**, Constitutive and regulated secretion, *Annu. Rev. Cell Biol.*, 3, 243, 1987.
57. **Kelly, R. B.**, Pathways of protein secretion in eukaryotes, *Science*, 230, 25, 1985.
58. **Klausner, R. D.**, Sorting and traffic in the central vacuolar system, *Cell*, 57, 703, 1989.
59. **Pelham, H. R. B.**, Control of protein exit from the endoplasmic reticulum, *Annu. Rev. Cell Biol.*, 5, 1, 1989.
60. **Rose, J. K. and Doms, R. W.**, Regulation of protein export from the endoplasmic reticulum, *Annu. Rev. Cell Biol.*, 4, 257, 1988.
61. **Rothman, J. E.**, Protein sorting by selective retention in the endoplasmic reticulum and Golgi stack, *Cell*, 50, 521, 1987.
62. **Dorel, C., Voelker, T. A., Herman, E. M., and Chrispeels, M. J.**, Transport of proteins to the plant vacuole is not by bulk flow through the secretory system, and requires positive sorting information, *J. Cell Biol.*, 108, 327, 1989.
63. **Hunt, D. and Chrispeels, M. J.**, The signal peptide of a vacuolar protein is necessary for the efficient secretion of a cytosolic protein, *Plant Physiol.*, 96, 18, 1991.
64. **Denecke, J., Botterman, J., and Deblaere, R.**, Protein secretion in plant cells can occur via a default pathway, Plant Cell, 2, 51, 1990.
65. **Iturriaga, G., Jefferson, R. A., and Bevan, M. V.**, Endoplasmic reticulum targeting and glycosylation of hybrid proteins in transgenic tobacco, *Plant Cell*, 1, 381, 1989.
66. **Wieland, F. T., Gleason, M. L., Serafini, T. A., and Rothman, J. E.**, The rate of bulk flow from the endoplasmic reticulum to the cell surface, *Cell*, 50, 289, 1987.
67. **Gongalez-Noriega, A., Grubb, J. H., Talkad, V., and Sly, W. S.**, Chloroquine inhibits lysosomal enzyme pinocytosis and enhances lysosomal enzyme secretion by impairing receptor recycling, *J. Cell Biol.*, 85, 839, 1980.
68. **Hasilik, A. and Neufeld, E. F.**, Biosynthesis of lysosomal enzymes in fibroblasts. Synthesis as precursors of higher molecular weight, *J. Biol. Chem.*, 255, 4937, 1980.
69. **Stevens, T. H., Rothman, J. H., Payne, G. S., and Schekman, R.**, Gene dosage-dependent secretion of yeast vacuolar carboxypeptidase Y, *J. Cell Biol.*, 102, 1551, 1986.
70. **Munro, S. and Pelham, H. R. B.**, A C-terminal signal prevents secretion of luminal ER proteins, *Cell*, 48, 988, 1987.
71. **Poruchynsky, M. S., Tyndall, C., Both, G. W., Sato, F., Bellamy, A. R., and Atkinson, P. A.**, Deletions into an NH_2-terminal hydrophobic domain result in secretion of rotovirus VP7, a resident membrane glycoprotein, *J. Cell Biol.*, 101, 2199, 1985.
72. **Pääbo, S., Bhat, B. M., Wold, W. S. M., and Peterson, P. A.**, A short sequence in the COOH-terminus makes an adenovirus membrane glycoprotein a resident of the endoplasmic reticulum, *Cell*, 50, 311, 1987.
73. **Hurtley, S. M. and Helenius, A.**, Protein oligomerization in the endoplasmic reticulum, *Annu. Rev. Cell Biol.*, 5, 277, 1989.
74. **Gething, M.-J. and Sambrook, J.**, Protein folding and intracellular transport: studies on influenza virus hemagglutinin, *Biochem. Soc. Symp.*, 55, 155, 1989.
75. **Moore, H.-P. H., Gumbiner, B., and Kelly, R. B.**, Chloroquine diverts ACTH from a regulated to a constitutive secretory pathway in AtT-20 cells, *Nature*, 302, 434, 1983.
76. **Wagner, D. D., Mayadas, T., and Marder, V. J.**, Initial glycosylation and acidic pH in the Golgi apparatus are required for the multimerization of von Willebrand factor, *J. Cell Biol.*, 102, 1320, 1986.
77. **Tartakoff, A. M.**, Perturbation of vesicular traffic with the carboxylic ionophore monensin, *Cell*, 32, 1026, 1983.
78. **Craig, S. and Goodchild, D. J.**, Golgi-mediated vicilin accumulation in pea cotyledon cells is redirected by monensin and nigericin, *Protoplasma*, 122, 91, 1984.
79. **Bowles, D. J., Marcus, S. E., Pappin, J. C., Findlay, J. B. C., Eliopoulos, E., Maylox, P. R., and Burgess, J.**, Post-translational processing of concanavalin A precursors in jackbean cotyledons, *J. Cell Biol.*, 102, 1284, 1986.
80. **Narváez-Vásquez, J., Franceschi, V. R., and Ryan, C. A.**, Proteinase-inhibitor synthesis in tomato plants: evidence for extracellular deposition in roots through the secretory pathway, *Planta*, 189, 257, 1993.
81. **Chrispeels, M. J.**, The Golgi apparatus mediates the transport of phytohemagglutinin to the protein bodies in bean cotyledons, *Planta*, 158, 140, 1983.

82. Wilkins, T. A., Bednarek, S. Y., and Raikhel, N. V., Role of propeptide glycan in post-translational processing and transport of barley lectin to vacuoles in transgenic tobacco, *Plant Cell*, 2, 301, 1990.
83. Bednarek, S. Y. and Raikhel, N. V., Intracellular trafficking of secretory proteins, *Plant Mol. Biol.*, 20, 133, 1992.
84. De Loose, M., Tire, G. G., Villarroel, R., Genetello, C., van Montagu, M., Depicker, A., and Inzé, D., The extensin signal peptide allows secretion of a heterologous protein from protoplasts, *Gene*, 99, 95, 1991.
85. Kornfeld, S. and Mellman, I., The biogenesis of lysosomes, *Annu. Rev. Cell Biol.*, 5, 483, 1989.
86. Farguhar, M. G., Progress in unravelling pathways of Golgi traffic, *Annu. Rev. Cell Biol.*, 1, 447, 1985.
87. Gebhart, A. M. and Ruddon, R. W., What regulates secretion of non-stored proteins by eukaryotic cells? *Bioessays*, 4, 213, 1986.
88. Palade, G. E., Intracellular aspects of the process of protein synthesis, *Science*, 189, 347, 1975.
89. Chung, K.-N., Walter, P., Aponte, G. W., and Moore, H.-P. H., Molecular sorting in the secretory pathway, *Science*, 243, 192, 1989.
90. Ho, D. T.-H., Hormonal and genetic regulation of α-amylase synthesis in barley aleurone cells, in *Genome Organization and Expression in Plants*, Leaver, C. J., Ed., Plenum Press, New York, 1980, 147.
91. Simon, P. and Jones, R. L., Synthesis and secretion of catalytically active barley α-amylase isoforms by *Xenopus* oocytes injected with barley mRNAs, *Eur. J. Cell Biol.*, 47, 213, 1988.
92. Buchanan, M. J., Iman, S. H., Eskue, W. A., and Snell, W. J., Activation of the cell wall degrading protease, lysin during sexual signalling in *Chlamydomonas*: the enzyme is stored as an inactive, higher relative molecular mass precursor in the periplasm, *J. Cell Biol.*, 108, 199, 1989.
93. Rapoport, T. A., Transport of proteins across the endoplasmic reticulum membrane, *Science*, 258, 931, 1992.
94. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D., *Molecular Biology of the Cell*, 3rd ed., Garland Publishing, New York, 1994.
95. von Heijne, G., Signal sequences. The limits of variation, *J. Mol. Biol.*, 184, 99, 1985.
96. von Heijne, G., Towards a comparative anatomy of N-terminal topogenic protein sequences, *J. Mol. Biol.*, 189, 239, 1986.
97. von Heijne, G., Transcending the impermeable: how proteins come to terms with membranes, *Biochim. Biophys. Acta*, 947, 307, 1988.
98. Bennett, A.B. and Osteryoung, K. W., Protein transport and targeting within the endomembrane system of plants, in *Plant Genetic Engineering*, Grierson, D., Ed., Chapman and Hall, New York, 1991, 199.
99. Walter, P., Gilmore, R., and Blobel, G., Protein translocation across the endoplasmic reticulum, *Cell*, 38, 5, 1984.
100. Walter, P. and Lingappa, V. R., Mechanism of protein translocation across the endoplasmic reticulum membrane, *Annu. Rev. Cell Biol.*, 2, 499, 1986.
101. Rothman, J. E., GTP and methionine bristles, *Nature*, 340, 433, 1989.
102. Gilmore, R., Protein translocation across the endoplasmic reticulum: a tunnel with toll booths at entry and exit, *Cell*, 75, 589, 1993.
103. Rothstein, S. J., Lazarus, C. M., Smith, W. E., Baulcombe, D. C., and Gatenby, A. A., Secretion of a wheat α-amylase expressed in yeast, *Nature*, 308, 662, 1984.
104. Rothstein, S. J., Lahners, K. N., Lazarus, C. M., Baulcombe, D. C., and Gatenby, A. A., Synthesis and secretion of wheat α-amylase in *Saccharomyces cerevisiae*, *Gene*, 55, 353, 1987.
105. Lindstrom, J. T., Chu, B., and Belanger, F. C., Isolation and characterization of an *Arabidopsis thaliana* gene for the 54 kDa subunit of the signal recognition particle, *Plant Mol. Biol.*, 23, 1265, 1993.
106. Lütcke, H., High, S., Römisch, K., Ashford, A. J., and Dobberstein, B., The methionine-rich domain of the 54 kDa subunit of signal recognition particle is sufficient for the interaction with signal sequences, *EMBO J.*, 11, 1543, 1992.
107. Deshaies, R. J., Koch, B. D., Werner-Washburne, M., Craig, E. A., and Schekman, R., A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides, *Nature*, 332, 800, 1988.
108. Görlich, D. and Rapoport, T. A., Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane, *Cell*, 75, 615, 1993.
109. Simon, S. and Blobel, G., A protein-conducting channel in the ER, *Cell*, 65, 371, 1991.
110. Crowley, K. S., Reinhart, G. D., and Johnson, A. E., The signal sequence moves through a ribosomal tunnel into a noncytoplasmic aqueous environment at the ER membrane early in translocation, *Cell*, 73, 1101, 1993.
111. Rothblatt, J. A., Deshaies, R. J., Sanders, S. L., Daum, G., and Schekman, R., Multiple genes are required for proper insertion of secretory proteins into the endoplasmic reticulum in yeast, *J. Cell Biol.*, 109, 2641, 1989.
112. Deshaies, R. J., Sanders, S. L., Feldheim, D. A., and Schekman, R., Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex, *Nature*, 349, 806, 1991.
113. Sanders, S. L., Whitfield, K. M., Vogel, J. P., Rose, M. D., and Schekman, R. W., Sec61p and BiP directly facilitate polypeptide translocation into the ER, *Cell*, 69, 353, 1992.
114. Görlich, D., Prehn, S., Hartmann, E., Kalies, K.-U., and Rapoport, T. A., A mammalian homolog of Sec61p and SecYp is associated with ribosomes and nascent polypeptides during translocation, *Cell*, 71, 489, 1992.

115. Müsch, A., Wiedmann, M., and Rapoport, T. A., Yeast Sec proteins interact with polypeptides traversing the endoplasmic reticulum membrane, *Cell*, 69, 343, 1992.
116. Tazawa, S., Unuma, M., Tondokoro, N., Asano, Y., Ohsumi, T., Ichimura, T., and Sugano, H., Identification of a membrane protein responsible for ribosome binding in rough microsomal membranes, *J. Biochem.*, 109, 89, 1991.
117. Ichimura, T., Ohsumi, T., Shindo, Y., Ohwada, T., Yagame, H., Momose, Y., Omata, S., and Sugano, H., Isolation and some properties of a 34-kDa membrane protein that may be essential for ribosome binding in rat liver rough microsomes, *FEBS Lett.*, 296, 7, 1992.
118. Savitz, A. J. and Meyer, D. I., 180-kDa ribosome receptor is essential for both ribosome binding and protein translocation, *J. Cell Biol.*, 120, 853, 1993.
119. Tajima, S., Lauffer, L., Rath, V. L., and Walter, P., The signal recognition particle receptor is a complex that contains two distinct polypeptide chains, *J. Cell Biol.*, 103, 1167, 1986.
120. Klappa, P., Mayinger, P., Pipkorn, R., Zimmermann, M., and Zimmermann, R., A microsomal protein is involved in ATP-dependent transport of presecretory proteins into mammalian microsomes, *EMBO J.*, 10, 2795, 1991.
121. Wiech, H., Stuart, R., and Zimmerman, R., Role of cytosolic factors in the transport of proteins across membranes, *Sem. Cell Biol.*, 1, 55, 1990.
122. Wiedmann, M., Kurzchalia, T. V., Hartmann, E., and Rapoport, T. A., A signal sequence receptor in the endoplasmic reticulum membrane, *Nature*, 328, 830, 1987.
123. Rothman, J. E., Polypeptide chain binding proteins: catalysts of protein folding and related processes in cells, *Cell*, 59, 591, 1989.
124. Chirico, W. J., Waters, M. G., and Blobel, G., 70 K heat shock related proteins stimulate protein translocation into microsomes, *Nature*, 332, 805, 1988.
125. Dickinson, C. D., Floener, L. A., Lilley, G. G., and Nielsen, N. C., Self-assembly of proglycinin and hybrid proglycinin synthesized *in vitro* from cDNA, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 5525, 1987.
126. Miyata, S. and Akazawa, T., α-Amylase biosynthesis: signal sequence prevents normal conversion of the unprocessed precursor molecule to the biologically active form, *Proc. Natl. Acad. Sci. U.S.A.*, 79, 7792, 1982.
127. Pelham, H. R. B., Hardwick, K. G., and Lewis, M. J., Sorting of soluble ER proteins in yeast, *EMBO J.*, 7, 1757, 1988.
128. Pelham, H. R. B., The retention signal for soluble proteins of the endoplasmic reticulum, *Trends Biochem. Sci.*, 15, 483, 1990.
129. Freedman, R. B., Native disulphide bond formation in protein biosynthesis: evidence for the role of protein disulphide isomerase, *Trends Biochem. Sci.*, 9, 438, 1984.
130. Geetha-Habib, M., Noiva, R., Kaplan, H. A., and Lennarz, W. J., Glycosylation site binding protein, a component of oligosaccharyl transferase, is highly similar to three other 57 kDa luminal proteins of the ER, *Cell*, 54, 1053, 1988.
131. Munro, S. and Pelham, H. R. B., An Hsp70-like protein in the ER: identity with the 78 kDa glucose-regulated protein and immunoglobulin heavy chain binding protein, *Cell*, 46, 291, 1986.
132. Tillmann, U., Viola, G., Kayser, B., Siemeister, G., Hesse, T., Palme, K., Lobler, M., and Klambt, D., cDNA clones of the auxin-binding protein from corn coleoptiles (*Zea mays* L.): isolation and characterization by immunogold methods, *EMBO J.*, 8, 2463, 1989.
133. Inhohara, N., Shimomura, S., Fukui, T., and Futai, M., Auxin-binding protein located in the endoplasmic reticulum of maize shoots: molecular cloning and complete primary structure, *Proc. Natl. Acad. Sci. U.S.A.*, 86, 3564, 1989.
134. Hesse, T., Feldwisch, J., Balshusemann, D., Bauw, G., Puype, M., Vandekerckhove, J., Lobler, M., Klambt, D., Schell, J., and Palme, K., Molecular cloning and structural analysis of a gene from *Zea mays* (L.) coding for a putative receptor for the plant hormone auxin, *EMBO J.*, 8, 2453, 1989.
135. Meyer, D. J. and Bennett A. B., Characterization of the ER localized molecular chaperone BiP in tomato, *Intl. Soc. Plant Mol. Biol.*, Third Intl. Congress, Abstract 990, 1991.
136. Fontes, E. B. P., Shank, B. B., Wrobel, R. L., Moose, S. P., O'Brian, G. R., Wurtzel, E. T., and Boston, R. S., Characterization of an immunoglobulin binding protein homologue in the maize floury-2 endosperm mutant, *Plant Cell*, 3, 483, 1991.
137. Boston, R. S., Fontes, E. B. P., Shank, B. B., and Wrobel, R. L., Increased expression of the maize immunoglobulin binding protein homologue b-70 in three zein regulatory mutants, *Plant Cell*, 3, 497, 1991.
138. Shorrosh, B. S. and Dixon, R. A., Molecular cloning of a putative endomembrane protein resembling vertebrate protein disulfide isomerase and a phosphotidylinositol-specific phospholipase C, *Proc. Natl. Acad. Sci. U.S.A.*, 88, 10941, 1991.
139. Denecke, J., Goldman, M. H. S., Demolder, J., Seurinck, J., and Botterman, J., The tobacco luminal binding protein is encoded by a multigene family, *Plant Cell*, 3, 1025, 1991.
140. Denecke, J., De Rycke, R., and Botterman, J., Plant and mammalian sorting signals for protein retention in the endoplasmic reticulum contain a conserved epitope, *EMBO J.*, 11, 2345, 1992.
141. Denecke, J., Ek, B., Caspers, M., Sinjorgo, K. M. C., and Palva, E. T., Analysis of sorting signals responsible for the accumulation of soluble reticuloplasmins

in the plant endoplasmic reticulum, *J. Exp. Bot.*. 44 (Suppl.), 213, 1993.

142. Napier, R. M., Fowke, L. C., Hawes, C., Lewis, M., and Pelham, H. R. B., Immunological evidence that plants use both HDEL and KDEL for targeting proteins to the endoplasmic reticulum, *J. Cell Sci.*, 102, 261, 1992.

143. Pelham, H. R. B., Evidence that luminal ER proteins are sorted from secreted proteins in a post-ER compartment, *EMBO J.*, 7, 913, 1988.

144. Mazzarella, R. A., Srinivasan, M., Haugejorden, S.M., and Green, M., ERp72, an abundant luminal endoplasmic reticulum protein contains three copies of the active site sequences of protein disulfide isomerase, *J. Biol. Chem.*, 265, 1094, 1990.

145. Ceriotti, A. and Colman, A., Binding to membrane proteins within the endoplasmic reticulum cannot explain the retention of the glucose-regulated protein GRP78 in *Xenopus* oocytes, *EMBO J.*, 7, 633, 1988.

146. Andres, D. A., Dickerson, I. M., and Dixon, J. E., Variants of the carboxy-terminal KDEL sequence direct intracellular retention, *J. Biol. Chem.*, 265, 5952, 1990.

147. Arber, S., Krause, K.-H., and Caroni, P., s-Cyclophilin is retained intracellularly via a unique COOH-terminal sequence and co-localizes with the calcium protein calreticulin, *J. Cell Biol.*, 116, 113, 1992.

148. Jones, A. M. and Herman, E. M., KDEL-containing auxin binding protein is secreted to the plasma membrane and cell wall, *Plant Physiol.*, 101, 595, 1993.

149. Napier, R. M. and Venis, M. A., Monoclonal antibodies detect an auxin-induced conformational change in the maize auxin-binding protein, *Planta*, 182, 313, 1990.

150. Yoshimori, T., Semba, T., Takemoto, H., Akagi, S., Yamamoto, A., and Tashiro, Y., Protein disulfide-isomerase in rat exocrine pancreatic cells is exported from the endoplasmic reticulum despite possessing the retention signal, *J. Biol. Chem.*, 265, 15984, 1990.

151. Pelham, H. R. B., Roberts, L. M., and Lord, J. M., Toxin entry. How reversible is the secretory pathway? *Trends Cell Biol.*, 2, 183, 1992.

152. Sixma, T. K., Pronk, S. E., Kalk, K. H., van Zanten, B. A. M., Berghuis, A. M., and Hol, W. G. J., Lactose binding to heat-labile enterotoxin revealed by X-ray crystallography, *Nature*, 355, 561, 1992.

153. Zagouras, P. and Rose, J. K., Carboxyterminal SEKDEL sequences retard but do not retain two secretory proteins in the endoplasmic reticulum, *J. Cell Biol.*, 109, 2633, 1989.

154. Herman, E. M., Tague, B. W., Hoffman, L. M., Kjemtrup, S. E., and Chrispeels, M. J., Retention of phytohemagglutinin with carboxyterminal KDEL in the nuclear envelope and the endoplasmic reticulum, *Planta*, 182, 305, 1990.

155. Vaux, D., Tooze, J., and Fuller, S., Identification by anti-idiotype antibodies of an intracellular membrane protein that recognizes a mammalian endoplasmic reticulum retention signal, *Nature*, 345, 495, 1990.

156. Wandelt, C. I., Khan, M. R. I., Craig, S., Schroeder, H. E., Spencer, D., and Higgins, T. J. V., Vicilin with carboxy-terminal KDEL is retained in the endoplasmic reticulum and accumulates to high levels in the leaves of transgenic plants, *Plant J.*, 2, 181, 1992.

157. Wandelt, C., Knibb, W., Schroeder, H. E., Khan, M. R. I., Spencer, D., Craig, S., and Higgins, T. J. V., The expression of an ovalbumin and a seed protein gene in the leaves of transgenic plants, in *Plant Molecular Biology, 1990*, Herrmann, R. G. and Larkins, B. A., Eds., Plenum Press, New York, 1991, 471.

158. Vaux, D., Tooze, J., and Fuller, S., Identification of an intracellular receptor for the KDEL retention signal, *J. Cell Biol.*, 109, 99, 1989.

159. Dean, N. and Pelham, H. R. B., Recycling of proteins from the Golgi compartment to the ER in yeast, *EMBO J.*, 9, 623, 1990.

160. Semenza, J. C., Hardwick, K. G., Dean, N., and Pelham, H. R. B., ERD2, a yeast gene required for the receptor-mediated retrieval of luminal ER proteins from the secretory pathway, *Cell*, 61, 1349, 1990.

161. Hardwick, K. G., Lewis, M. J., Semenza, J., Dean, N., and Pelham, H. R. B., ERD1, a yeast gene required for the retention of luminal endoplasmic reticulum proteins, affects glycoprotein processing in the Golgi apparatus, *EMBO J.*, 9, 623, 1990.

162. Lewis, M. J., Sweet, D. J., and Pelham, H. R. B., The ERD2 gene determines the specificity of the luminal ER protein retention system, *Cell*, 61, 1359, 1990.

163. Lewis, M. and Pelham, H. R. B., A human homologue of the yeast HDEL receptor, *Nature*, 348, 162, 1990.

164. Lewis, M. and Pelham, H. R. B., Ligand induced redistribution of a human KDEL receptor from the Golgi complex to the ER, *Cell*, 68, 353, 1992.

165. Tang, B. L., Wong, S. H., Qi, X. L., Low, S. H., and Hong, W., Molecular cloning, characterization, subcellular localization and dynamics of p23, the mammalian KDEL receptor, *J. Cell Biol.*, 120, 325, 1993.

166. Hsu, V. W., Shah, N., and Klausner, R. D., A Brefeldin A-like phenotype is induced by the overexpression of a human ERD2-like protein, ELP-1, *Cell*, 69, 625, 1992.

167. Lee, H.-I., Gal, S., Newman, T. C., and Raikhel, N. V., The *Arabidopsis* endoplasmic reticulum retention receptor functions in yeast, *Proc. Natl. Acad. Sci. U.S.A.*, 90, 11433, 1993.

168. Denecke et al., unpublished; cited in Reference 4.

169. Lippincott-Schwartz, J., Donaldson, J. G., Schweizer, A., Berger, E. G., Hauri, H.-P., Yuan, L. C., and Klausner, R. D., Microtubule-dependent retrograde transport of proteins into the ER in the presence of Brefeldin A suggests an ER recycling pathway, *Cell*, 60, 821, 1990.

170. Armstrong, J. and Warren, G., Violating the one-way system, *Nature*, 344, 383, 1990.

171. Lee, C. and Chen, L. B., Dynamic behavior of endoplasmic reticulum in living cells, *Cell*, 54, 37, 1988.

172. Klausner, R. B., Donaldson, J. G., and Lippincott-Schwartz, J., Brefeldin A: insights into the control of membrane traffic and organelle structure, *J. Cell Biol.*, 116, 1071, 1992.

173. Pelham, H. R. B., Multiple targets for brefeldin A, *Cell*, 67, 449, 1991.

174. Doms, R. W., Russ, G., and Yewdell, J. W., Brefeldin A redistributes resident and itinerant Golgi proteins to the endoplasmic reticulum, *J. Cell Biol.*, 109, 51, 1989.

175. Nakona, A., Brada, D., and Schekman, R., A membrane glycoprotein, Sec12p, required for protein transport from the endoplasmic reticulum to the Golgi apparatus in yeast, *J. Cell Biol.*, 107, 851, 1988.

176. Robinson, D. G., Brefeldin A: a tool for plant cell biologists? *Bot. Acta*, 106, 107, 1993.

177. Satiat-Jeunemaitre, B. and Hawes, C., G.A.T.T. (a general agreement on traffic and transport) and brefeldin A in plant cells, *Plant Cell*, 6, 463, 1994.

178. Holwerda, B. C., Padgett, H. S., and Rogers, J. C., Proaleurain vacuolar targeting is mediated by short contiguous peptide interactions, *Plant Cell*, 4, 307, 1992.

179. Driouich, A., Zhang, G. F., and Staehelin, L. A., Effect of brefeldin A on the structure of the Golgi apparatus and on the synthesis and secretion of proteins and polysaccharides in sycamore maple (*Acer pseudoplatanus*) suspension-cultured cells, *Plant Physiol.*, 101, 1363, 1993.

180. Gomez, L. and Chrispeels, M. J., Tonoplast and soluble vacuolar proteins are targeted by different mechanisms, *Plant Cell*, 5, 1113, 1993.

181. Satiat-Jeunemaitre, B. and Hawes, C., Redistribution of a Golgi glycoprotein in plant cells treated with brefeldin A, *J. Cell Sci.*, 103, 1153, 1992.

182. Sweet, D. J. and Pelham, H. R. B., The *Saccharomyces cerevisiae* SEC20 gene encodes a membrane glycoprotein which is sorted by the HDEL retrieval system, *EMBO J.*, 11, 423, 1992.

183. Jackson, M., Nilsson, T., and Peterson, P., Retrieval of transmembrane proteins to the endoplasmic-reticulum, *J. Cell Biol.*, 121, 317, 1993.

184. Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H. M., and Somerville, C. R., Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*, *Science*, 258, 1353, 1992.

185. Nilsson, T., Jackson, M., and Peterson, P. A., Short cytoplasmic sequences serve as retention signals for transmembrane proteins in the endoplasmic reticulum, *Cell*, 58, 707, 1989.

186. Wada, I., Rindress, D., Cameron, P., Ou, W.-J., Doherty II, J. J., Louvard, D., Bell, A. W., Dignard, D., Thomas, D. Y., and Bergeron, J. J. M., SSRα and associated calnexin are major calcium binding proteins of the endoplasmic reticulum membrane, *J. Biol. Chem.*, 266, 19599, 1991.

187. Mallabiabarrena, A., Fresno, M., and Alarcon B., An endoplasmic reticulum retention signal in the CD3ε chain of the T-cell receptor, *Nature*, 357, 593, 1992.

188. Stirzaker, S. C. and Both, G. W., The signal peptide of the rotovirus glycoprotein VP7 is essential for its retention in the endoplasmic reticulum as an integral membrane protein, *Cell*, 56, 741, 1989.

189. Kuroki, K., Russnak, R., and Ganem, D., Novel N-terminal amino acid sequence required for retention of a hepatitis B virus glycoprotein in the endoplasmic reticulum, *Mol. Cell. Biol.*, 9, 4459, 1989.

190. Brands, R., Snider, M. D., Hino, Y., Park, S. S., Gelboin, H. V., and Rothman, J. E., Retention of membrane proteins by the endoplasmic reticulum, *J. Cell Biol.*, 101, 1724, 1985.

191. Yamamoto, A., Masaki, R., and Tashiro, Y., Is cytochrome P-450 transported from the endoplasmic reticulum to the Golgi apparatus in rat hepatocytes? *J. Cell Biol.*, 101, 1733, 1985.

192. Schindler, R., Itin, C., Zerial, M., Lottspeich, F., and Hauri, H.-P., ERGIC-53, a membrane protein of the ER-Golgi intermediate compartment, carries an ER retention motif, *Eur. J. Cell Biol.*, 61, 1, 1993.

193. Nilsson, T., Lucocq, J. M., Mackay, D., and Warren, G., The membrane spanning domain of β-1,4-galactosyl-transferase specifies *trans*-Golgi localization, *EMBO J.*, 10, 3567, 1991.

194. Nilsson, T., Hoe, M. H., Slusarewicz, P., Rabouille, C., Watson, R., Hunte, F., Watzele, G., Berger, E. G., and Warren, G., Kin recognition between *medial* Golgi enzymes in HeLa cells, *EMBO J.*, 13, 562, 1994.

195. Burke, J., Pettitt, J. M., Schachter, H., Sarkar, M., and Gleeson, P. A., The transmembrane and flanking sequences of β1,2-*N*-acetylglucosaminyl-transferase I specify *medial*-Golgi localization, *J. Biol. Chem.*, 267, 24433, 1992.

196. Ponnambalam, S., Rabouille, C., Luzio, J. P., Nilsson, T., and Warren, G., The TGN38 glycoprotein contains two non-overlapping signals that mediate localization to the *trans*-Golgi network, *J. Cell Biol.*, 125, 253, 1994.

197. Bretscher, M. S. and Munro, S., Cholesterol and the Golgi apparatus, *Science*, 261, 1280, 1993.

198. Munro, S., Sequences within and adjacent to the transmembrane segment of α-2,6-sialyltransferase specify Golgi retention, *EMBO J.*, 10, 3577, 1991.

199. Harper, J. F., Surowy, T. K., and Sussman, M. R., Molecular cloning and sequence of cDNA encoding the plasma membrane proton pump (H^+-ATPase) of *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. U.S.A.*, 86, 1234, 1989.

200. Johnson, K. D., Höfte, H., and Chrispeels, M. J., An intrinsic tonoplast protein of protein storage vacu-

oles in seeds is structurally related to a bacterial solute transporter (Glpf), *Plant Cell*, 2, 525, 1990.
201. **Rea, P. A., Kim, Y., Sarafian, V., Poole, R. J., Davies, J. M., and Sanders, D.**, Vacuolar H⁺-translocating pyrophosphatases: a new category of ion translocase, *Trends Biochem. Sci.*, 17, 348, 1992.
202. **Sarafian, V., Kim, Y., Poole, R. J., and Rea, P. A.**, Molecular cloning and sequence of cDNA encoding the pyrophosphate-energized vacuolar membrane proton pump of *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. U.S.A.*, 89, 1775, 1992.
203. **Höfte, H. and Chrispeels, M. J.**, Protein sorting to the vacuolar membrane, *Plant Cell*, 4, 995, 1992.
204. **Melroy, D. L. and Herman, E. M.**, TIP, an integral membrane protein of the protein-storage vacuoles of the soybean cotyledon, undergoes developmentally regulated membrane accumulation and removal, *Planta*, 184, 113, 1991.
205. **Sze, H., Ward, J. M., and Lai, S.**, Vacuolar H⁺-translocating ATPase from plants: structure, function and isoforms, *J. Bioenerg. Biomemb.*, 24, 371, 1992.
206. **Maathuis, F. J. M. and Sanders, D.**, *Curr. Opin. Cell Biol.*, 4, 661, 1992.
207. **Martinoia, E.**, Transport processes in vacuoles of higher plants, *Bot. Acta*, 105, 232, 1992.
208. **Wirtz, K. W. A. and Gadella Jr., T. W. J.**, Properties and modes of action of specific and non-specific phospholipid transfer proteins, *Experientia*, 46, 592, 1990.
209. **Bankaitis, V. A., Aitken, J. R., Cleves, A. E., and Dowhan, W.**, An essential role for a phospholipid transfer protein in yeast Golgi function, *Nature*, 347, 561, 1990.
210. **Thoma, S., Kaneko, Y., and Somerville, C.**, A non-specific lipid transfer protein from *Arabidopsis* is a cell wall protein, *Plant J.*, 3, 427, 1993.
211. **Thoma, et al.**, unpublished; cited in Reference 50.
212. **Gardiner, M. and Chrispeels, M. J.**, Involvement of the Golgi apparatus in the synthesis and secretion of hydroxyproline-rich cell wall glycoproteins, *Plant Physiol.*, 55, 536, 1975.
213. **Driouich, A., Faye, L., and Staehelin, L. A.**, The plant Golgi apparatus: a factory for complex polysaccharides and glycoproteins, *Trends Biochem. Sci.*, 18, 210, 1993.
214. **Stafstrom, J. P. and Staehelin, L. A.**, The role of carbohydrate in maintaining extensin in an extended conformation, *Plant Physiol.*, 81, 242, 1986.
215. **Jentoft, N.**, Why are proteins O-glycosylated? *Trends Biochem. Sci.*, 15, 291, 1990.
216. **Kaushal, G. P. and Elbein, A. D.**, Glycoprotein processing enzymes of plants, *Methods Enzymol.*, 179, 452, 1989.
217. **Faye, L., Johnson, K. D., Sturm, A., and Chrispeels, M. J.**, Structure, biosynthesis, and function of asparagine-linked glycans on plant glycoproteins, *Physiol. Plant.*, 75, 309, 1989.
218. **Sturm, A., van Kuik, J. A., Vliegenthart, J. F. G., and Chrispeels, M. J.**, Structure, position, and biosynthesis of the high mannose and the complex oligosaccharide side chains of the bean storage protein phaseolin, *J. Biol. Chem.*, 262, 13392, 1987.
219. **Bewley, J. D. and Black, M.**, *Seeds. Physiology of Development and Germination*, 2nd ed., Plenum Press, New York, 1994.
220. **Elbein, A. D.**, The role of the lipid-linked saccharides in the biosynthesis of complex carbohydrates, *Annu. Rev. Plant Physiol.*, 30, 239, 1979.
221. **Hirschberg, C. B. and Snider, M. D.**, Topography of glycosylation in the rough endoplasmic reticulum and Golgi apparatus, *Annu. Rev. Biochem.*, 56, 63, 1987.
222. **Sturm, A., Johnson, K. D., Szumilo, T., Elbein, A. D., and Chrispeels, M. J.**, Subcellular localization of glycosidases and glycosyl-transferases involved in the processing of N-linked oligosaccharides, *Plant Physiol.*, 85, 741, 1987.
223. **Sturm, A., Chrispeels, M. J., Wieruszeski, J. M., Strecker, G., and Montreuil, J.**, Structural analysis of the N-linked oligosaccharides from jackbean α-mannosidase, *Glycoconjugates, Proc. 9th Intl. Symp.*, B28, 1987.
224. **Abeijon, C. and Hirschberg, C. B.**, Topography of glycosylation reactions in the endoplasmic reticulum, *Trends Biochem. Sci.*, 17, 32, 1992.
225. **Drake Jr., R. R., Kaushal, G. P., Pastuszak, I., and Elbein, A. D.**, Partial purification, photoaffinity labelling, and properties of mung bean UDP-glucose: dolicholphosphate glucosyl-transferase, *Plant Physiol.*, 97, 396, 1991.
226. **Kelleher, D. J., Kreibich, G., and Gilmore, R.**, Oligosaccharyl transferase activity is associated with a protein complex composed of ribophorins I and II and a 48 kDa protein, *Cell*, 69, 55, 1992.
227. **Bollini, R., Vitale, A., and Chrispeels, M. J.**, *In vivo* and *in vitro* processing of seed reserve protein in the endoplasmic reticulum: evidence for two glycosylation steps, *J. Cell Biol.*, 96, 999, 1983.
228. **Cramer, J. H., Lea, K., and Slightom, J. L.**, Expression of phaseolin cDNA genes in yeast under control of natural plant DNA sequences, *Proc. Natl. Acad. Sci. U.S.A.*, 82, 334, 1985.
229. **Ceriotti, A., Pedrazzini, E., Fabbrini, M. S., Zoppè, M., Bollini, R., and Vitale, A.**, Expression of wild-type and mutated vacuolar storage protein phaseolin in *Xenopus* oocytes reveals relationships between assembly and intracellular transport, *Eur. J. Biochem.*, 202, 959, 1991.
230. **Pedrazzini, et al.**, unpublished; cited in Reference 4.
231. **Bulleid, N. J., Bassel-Duby, R. S., Freedman, R. B., Sambrook, J. F., and Gething, M.-J. H.**, Cell-free synthesis of enzymically active tissue-type plasminogen activator. Protein folding determines the extent of N-linked glycosylation, *Biochem. J.*, 286, 275, 1992.
232. **Faye, L. and Chrispeels, M. J.**, Common antigenic determinants in the glycoproteins of plants, molluscs, and insects, *Glycoconjugate J.*, 5, 245, 1988.
233. **Faye, L., et al.**, in *Post Translational Modification and its Significance in Plant Development*, Battey, N. M., Dikinson, H. G., and Heatherington, A. M., Eds., Cambridge University Press, 1992, 213.
234. **Ali, M. S., Mitsui, T., and Akazawa, T.**, Golgi-specific localization of transglycosylases engaged in glycoprotein biosynthesis in suspension-cultured cells of sycamore (*Acer pseudoplatanus* L.), *Arch. Biochem. Biophys.*, 251, 421, 1986.
235. **Zhang, G. F. and Staehelin, L. A.**, Functional compartmentalization of the Golgi apparatus of plant cells: an immunocytochemical analysis of high pressure frozen and freeze-substituted sycamore maple suspension culture cells, *Plant Physiol.*, 99, 1070, 1992.
236. **Lainé, A. C., Gomord, V., and Faye, L.**, Xylose specific antibodies as markers of subcompartmentation of terminal glycosylation in the Golgi apparatus of sycamore cells, *FEBS Lett.*, 295, 179, 1991.
237. **Faye, L., Sturm, A., Bollini, R., Vitale, A., and Chrispeels, M. J.**, The position of the oligosaccharide side-chains of phytohemagglutinin and their accessibility to glycosidases determines their subsequent processing in the Golgi, *Eur. J. Biochem.*, 158, 655, 1986.
238. **Vitale, A., Warner, T. G., and Chrispeels, M. J.**, *Phaseolus vulgaris* phytohemagglutinin contains high-mannose and modified oligosaccharide chains, *Planta*, 160, 256, 1984.
239. **Sturm, A.**, Heterogeneity of the complex N-linked oligosaccharides at specific glycosylation sites of 2 secreted carrot glycoproteins, *Eur. J. Biochem.*, 199, 169, 1991.
240. **Woodward, J. R., Craik, D., Dell, A., Khoo, K.-H., Munro, S. L. A., Clarke, A. E., and Bacic, A.**, Structural analysis of the N-linked glycan chains from a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*, *Glycobiology*, 2, 241, 1992.
241. **Takahashi, N., Hotta, T., Ishihara, H., Mori, M., Tejima, S., Bligny, R., Akazawa, T., Endo, S., and Arata, Y.**, Xylose-containing common structural unit in N-linked oligosaccharides of laccase from sycamore cells, *Biochemistry*, 25, 388, 1986.
242. **Hayashi, M., Tsuru, A., Mitsui, T., Takahashi, N., Hanzawa, H., Arata, Y., and Akazawa, T.**, Structure and biosynthesis of the xylose-containing carbohydrate moiety of rice α-amylase, *Eur. J. Biochem.*, 191, 287, 1990.
243. **Vitale, A. and Chrispeels, M. J.**, Transient N-acetylglucosamine in the biosynthesis of phytohemagglutinin: attachment in the Golgi apparatus and removal in protein bodies, *J. Cell Biol.*, 99, 133, 1984.
244. **Herman, E. M., Shannon, L. M., and Chrispeels, M. J.**, Concanavalin A is synthesized as a glycoprotein precursor, *Planta*, 165, 23, 1985.
245. **Mansfield, M. A., Peumans, W. J., and Raikhel, N. V.**, Wheat-germ agglutinin is synthesized as a glycosylated precursor, *Planta*, 173, 482, 1988.
246. **Shinshi, H., Wenzler, H., Neuhaus, J., Felix, G., Hofsteenge, J. F., and Meins, J.**, Evidence for N- and C-terminal processing of a plant defense-related enzyme: primary structure of tobacco prepro-β-1,3-glucanase, *Proc. Natl. Acad. Sci. U.S.A.*, 85, 5541, 1988.
247. **Duksin, D. and Mahoney, W. C.**, Relationship of the structure and biological activity of the natural homologues of tunicamycin, *J. Biol. Chem.*, 257, 3105, 1982.
248. **Chrispeels, M. J., Higgins, T. J. V., Craig, S., and Spencer, D.**, Role of the endoplasmic reticulum in the synthesis of reserve proteins and the kinetics of their transport to protein bodies in developing pea cotyledons, *J. Cell Biol.*, 93, 5, 1982.
249. **Bollini, R., Ceriotti, A., Daminati, M. G., and Vitale, A.**, Glycosylation is not needed for the intracellular transport of phytohemagglutinin in developing *Phaseolus vulgaris* cotyledons and the maintenance of its biological activities, *Physiol. Plant.*, 65, 15, 1985.
250. **Sonnewald, U., von Schaewen, A., and Willmitzer, L.**, Expression of mutant patatin protein in transgenic tobacco: role of glycans and intracellular location, *Plant Cell*, 2, 345, 1990.
251. **Voelker, T. A., Herman, E. M., and Chrispeels, M. J.**, *In vitro* mutated phytohemagglutinin genes expressed in tobacco seeds: role of glycans in protein targeting and stability, *Plant Cell*, 1, 95, 1989.
252. **Schwaiger, H., Hasilik, A., von Figura, K., Wiemken, A., and Tanner, W.**, Carbohydrate-free carboxypeptidase is transferred into the lysosome-like yeast vacuole, *Biochem. Biophys. Res. Commun.*, 104, 950, 1982.
253. **Vitale, A., Zoppè, M., and Bollini, R.**, Mannose analog 1-deoxymannojirimycin inhibits the Golgi-mediated processing of bean storage glycoproteins, *Plant Physiol.*, 89, 1079, 1989.
254. **Elbein, A.D.**, Inhibitors of the biosynthesis and processing of N-linked oligosaccharide chains, *Annu. Rev. Biochem.*, 56, 497, 1987.
255. **Driouich, A., Gonnet, P., Makkie, M., Laine, A.-C., and Faye, L.**, The role of high-mannose and complex asparagine-linked glycans in the secretion and stability of glycoproteins, *Planta*, 180, 96, 1989.
256. **von Schaewen, A., Sturm, A., O'Neill, J., and Chrispeels, M. J.**, Isolation of a mutant *Arabidopsis* plant that lacks N-acetyl glucosaminyl transferase I and is unable to synthesize Golgi-modified complex N-linked glycans, *Plant Physiol.*, 102, 1109, 1993.
257. **Hori, H. and Elbein, A. D.**, Tunicamycin inhibits protein glycosylation in suspension-cultured soybean cells, *Plant Physiol.*, 67, 882, 1981.
258. **Ravi, K., Hu, C., Reddi, P. S., and Huystee, R. B. V.**, Effect of tunicamycin on peroxidase release by cultured peanut suspension cells, *J. Exp. Bot.*, 37, 1708, 1986.
259. **Faye, L. and Chrispeels, M. J.**, Apparent inhibition of β-fructosidase secretion by tunicamycin may be explained by breakdown of the unglycosylated protein during secretion, *Plant Physiol.*, 89, 845, 1989.

260. Bustos, M. M., Kalkan, F. A., Van den Bosch, K. A., and Hall, T. C., Differential accumulation of four phaseolin glycoforms in transgenic tobacco, *Plant Mol. Biol.*, 16, 381, 1991.
261. Nagai, K., Shibata, K., and Yamaguchi, H., Role of intramolecular high-mannose chains in the folding and assembly of soybean (*Glycine max*) lectin polypeptides: studies by the combined use of spectroscopy and gel-filtration size analysis, *J. Biochem.*, 114, 830, 1993.
262. De Vries, S. C., Booij, H., Janssens, R., Vogels, R., Saris, L., LoSchiavo, F., Terzi, M., and Van Kammen, A., Carrot somatic embryogenesis depends on the phytohormone-controlled presence of correctly glycosylated extracellular proteins, *Genes Dev.*, 2, 462, 1988.
263. Hammond, C., Braakman, I., and Helenius, A., Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control, *Proc. Natl. Acad. Sci. U.S.A.*, 91, 913, 1994.
264. Moore, S. E. and Spiro, R. G., Inhibition of glucose trimming by castanospermine results in rapid degradation of unassembled major histocompatibility complex Class I molecules, *J. Biol. Chem.*, 268, 3809, 1993.
265. Elbein, A. D., Glycosidase inhibitors: inhibitors of N-linked oligosaccharide processing, *FASEB J.*, 5, 3055, 1991.
266. Dickinson, C. D., Hussein, E. H. A., and Nielsen, N. C., Role of post-translational cleavage in glycinin assembly, *Plant Cell*, 1, 459, 1989.
267. Chrispeels, M. J. and Bollini, R., Characteristics of membrane-bound lectin in developing *Phaseolus vulgaris* cotyledons, *Plant Physiol.*, 70, 1425, 1982.
268. Van der Wilden, W. and Chrispeels, M. J., Characterization of the isozymes of α-mannosidase located in the cell wall, protein bodies, and endoplasmic reticulum of *Phaseolus vulgaris* cotyledons, *Plant Physiol.*, 71, 82, 1983.
269. Gething, M.-J. and Sambrook, J., Protein folding in the cell, *Nature*, 355, 33, 1992.
270. Georgopoulos, C., The emergence of the chaperone machines, *Trends Biochem. Sci.*, 17, 295, 1992.
271. Hartl, F.-U., Hlodan, R., and Langer, T., Molecular chaperones in protein folding: the art of avoiding sticky situations, *Trends Biochem. Sci.*, 19, 20, 1994.
272. D'Amico, L., Valsasina, B., Daminati, M. G., Fabbrini, M. S., Nitti, G., Bollini, R., Ceriotti, A., and Vitale, A., Bean homologues of the mammalian glucose regulated proteins: induction by tunicamycin and interaction with newly synthesized storage proteins in the endoplasmic reticulum, *Plant J.*, 2, 443, 1992.
273. Jones, R. L. and Bush, D. S., Gibberellic acid regulates the level of a BiP cognate in the endoplasmic reticulum of barley aleurone cells, *Plant Physiol.*, 97, 456, 1991.
274. Anderson, J. V., Neven, L. G., Li, Q.-B., Haskell, D. W., and Guy, C. L., A cDNA encoding the endoplasmic reticulum-luminal heat-shock protein from spinach (*Spinacia oleracea* L.), *Plant Physiol.*, 104, 303, 1994.
275. Shiu, R. P. C., Pouyssegur, J., and Pastan, I., Glucose depletion accounts for the induction of two transformation-sensitive membrane proteins in Rous sarcoma virus-transformed chick embryo fibroblasts, *Proc. Natl. Acad. Sci. U.S.A.*, 74, 3840, 1977.
276. Olden, K., Pratt, R. M., Jaworski, C. and Yamada, K. M., Evidence for role of glycoprotein carbohydrates in membrane transport: specific inhibition by tunicamycin, *Proc. Natl. Acad. Sci. U.S.A.*, 76, 791, 1979.
277. Mazzarella, R. A. and Green, M., ERp99, an abundant, conserved glycoprotein of the endoplasmic reticulum, is homologous to the 90 kDa heat shock protein (hsp90) and the 94 kDa glucose regulated protein (GRP94), *J. Biol. Chem.*, 262, 8875, 1987.
278. Bergeron, J. J. M., Brenner, M. B., Thomas, D. Y., and Williams, D. B., Calnexin: a membrane-bound chaperone of the endoplasmic reticulum, *Trends Biochem. Sci.*, 19, 124, 1994.
279. Huang, L., Franklin, A. E., and Hoffman, N. E., Primary structure and characterization of an *Arabidopsis thaliana* calnexin-like protein, *J. Biol. Chem.*, 268, 6560, 1993.
280. Lee, A. S., Coordinated regulation of a set of genes by glucose and calcium ionophores in mammalian cells, *Trends Biochem. Sci.*, 12, 20, 1987.
281. Kozutsumi, Y., Segal, M., Normington, K., Gething, M.-J., and Sambrook, J., The presence of malfolded proteins in the endoplasmic reticulum signals the induction of glucose-regulated proteins, *Nature*, 332, 462, 1988.
282. Kassenbrock, C. K., Garcia, P. D., Walter, P., and Kelly, R. B., Heavy-chain binding protein recognizes aberrant polypeptides translocated *in vitro*, *Nature*, 333, 90, 1988.
283. Rizzolo, L. J., Finidori, J., Gonzalez, A., Arpin, M., Ivanov, I. E., Adesnik, M., and Sabatini, D. D., Biosynthesis and intracellular sorting of growth hormone-viral envelope glycoprotein hybrids, *J. Cell Biol.*, 101, 1351, 1985.
284. Sharma, S., Rogers, L., Brandsma, J., Gething, M.-J., and Sambrook, J., SV40 T antigen and the exocytic pathway, *EMBO J.*, 4, 1479, 1985.
285. Pelham, H. R. B., Speculations on the functions of the major heat shock and glucose regulated proteins, *Cell*, 46, 959, 1986.
286. Degen, E., Cohen-Doyle, M. F., and Williams, D. B., *J. Exp. Med.*, 175, 1653, 1992.
287. Bole, D. G., Hendershot, L. M., and Kearny, J. F., Post-translational association of immunoglobulin heavy chain binding protein with nascent heavy chains in non-secreting and secreting hybridomas, *J. Cell Biol.*, 102, 1558, 1986.

288. Knittler, M. R. and Haas, I. G., Interaction of BiP with newly-synthesized immunoglobulin light chain molecules: cycles of sequential binding and release, *EMBO J.*, 11, 1573, 1992.
289. Rose, M. D., Misra, L. M., and Vogel, J. P., KAR2, a karyogamy gene, is the yeast homologue of the mammalian BiP/GRP 78 gene, *Cell*, 57, 1211, 1989.
290. Pedrazzini, E., Giovinazzo, G., Bollini, R., Ceriotti, A., and Vitale, A., The binding of BiP to an assembly-defective protein in plant cells, *Plant J.*, 5, 103, 1994.
291. Kohno, K., Normington, K., Sambrook, J., Gething, M.-J., and Mori, K., The promoter region of the yeast KAR2 (BiP) gene contains a regulatory domain that responds to the presence of unfolded proteins in the endoplasmic reticulum, *Mol. Cell. Biol.*, 13, 877, 1993.
292. Resendez, E., Jr., Wooden, S. K., and Lee, A. S., Identification of highly conserved regulatory domains and protein-binding sites in the promoters of the rat and human genes encoding the stress-inducible 78-kilodalton glucose-regulated protein, *Mol. Cell. Biol.*, 8, 4579, 1988.
293. Chang, S. C., Erwin, A. E., and Lee, A. S., Glucose-regulated protein (*GRP94* and *GRP78*) genes share common regulatory domains and are coordinately regulated by common *trans*-acting factors, *Mol. Cell. Biol.*, 9, 2153, 1989.
294. Walther-Larsen, H., Brandt, J., Collinge, D. B., and Thordal-Christensen, H., A pathogen-induced gene of barley encodes a HSP90 homologue showing striking similarity to vertebrate forms resident in the endoplasmic reticulum, *Plant Mol. Biol.*, 21, 1097, 1993.
295. Blond-Elguindi, S., Cwirla, S. E., Dower, W. J., Lipshutz, R. J., Sprang, S. R., Sambrook, J. F., and Gething, M.-J. H., Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP, *Cell*, 75, 717, 1993.
296. Flynn, G. C., Chappell, T. G., and Rothman, J. E., Peptide binding and release by proteins implicated as catalysts of protein assembly, *Science*, 245, 385, 1989.
297. Flynn, G. C., Pohl, J., Flocco, M. T., and Rothman, J. E., Peptide-binding specificity of the molecular chaperone BiP, *Nature*, 353, 726, 1991.
298. Macer, D. R. J. and Koch, G. L. E., Identification of a set of calcium binding proteins in reticuloplasm, the luminal content of the endoplasmic reticulum, *J. Cell Sci.*, 92, 61, 1988.
299. Ohsako, S., Hayashi, Y., and Bunick, D., Molecular cloning and sequencing of calnexin-t. An abundant male germ cell-specific calcium-binding protein of the endoplasmic reticulum, *J. Biol. Chem.*, 269, 14140, 1994.
300. Suzuki, S. K., Bonifacino, J. S., Lin, A. Y., Davis, M. M., and Klausner, R. D., Regulating the retention of T-cell receptor α chain variant within the endoplasmic reticulum: Ca^{2+}-dependent association with BiP, *J. Cell Biol.*, 114, 189, 1991.
301. Sambrook, J. F., The involvement of calcium in transport of secretory proteins from the endoplasmic reticulum, *Cell*, 61, 197, 1990.
302. Booth, C. and Koch, G. L. E., Perturbation of cellular calcium induces secretion of luminal ER proteins, *Cell*, 59, 729, 1989.
303. Clarke, S., Protein isoprenylation and methylation at carboxyl-terminal cysteine residues, *Annu. Rev. Biochem.*, 61, 355, 1992.
304. Rine, J. and Kim, S.-H., A role for isoprenoid lipids in the localization and function of an oncoprotein, *New Biol.*, 2, 219, 1990.
305. Black, S. D., Development of hydrophobicity parameters for prenylated proteins, *Biochem. Biophys. Res. Commun.*, 186, 1437, 1992.
306. Zhu, J.-K., Bressan, R. A., and Hasegawa, P. M., Isoprenylation of the plant molecular chaperone ANJ1 facilitates membrane association and function at high temperature, *Proc. Natl. Acad. Sci. U.S.A.*, 90, 8557, 1993.
307. Klausner, R. D. and Sitia, R., Protein degradation in the endoplasmic reticulum, *Cell*, 62, 611, 1990.
308. Dice, J. F., Peptide sequences that target cytosolic proteins for lysosomal proteolysis, *Trends Biochem. Sci.*, 15, 305, 1990.
309. Chen, C., Bonifacino, J. S., Yuan, L. C., and Klausner, R. D., Selective degradation of T-cell antigen receptor chains in a pre-Golgi compartment, *J. Cell Biol.*, 107, 2149, 1988.
310. Lippincott-Swartz, J., Bonifacino, J. S., Yuan, L. C., and Klausner, R. D., Degradation from the endoplasmic reticulum: disposing of newly synthesized proteins, *Cell*, 54, 209, 1988.
311. Hoffman, L. M., Donaldson, D. D., and Herman, E. M., A modified storage protein is synthesized, processed, and degraded in the seeds of transgenic plants, *Plant Mol. Biol.*, 11, 717, 1988.
312. Herman, E. M., Pueyo, J. J., and Chrispeels, M. J., Quality control of a seed storage protein with a destabilizing epitope occurs in the post-ER secretory system, *Plant Physiol.*, 105 (Suppl.), 94, 1994.
313. Freedman, R. B., Bulleid, N. J., Hawkins, H. C., and Paver, J. L., Role of protein disulphide-isomerase in the expression of native proteins, *Biochem. Soc. Symp.*, 55, 167, 1989.
314. Edman, J. C., Ellis, L., Blacher, R. W., Roth, R. A., and Rutter, W. J., Sequence of protein disulphide isomerase and implications of its relationship to thioredoxin, *Nature*, 317, 267, 1985.
315. Bulleid, N. J. and Freedman, R. B., Defective co-translational formation of disulphide bonds in protein disulphide-isomerase-deficient microsomes, *Nature*, 335, 649, 1988.
316. Wetterau, J. R., Combs, K. A., Spinner, S. N., and Joiner, B. G., Protein disulfide isomerase is

a component of the microsomal triglyceride transfer protein complex, *J. Biol. Chem.*, 265, 9800, 1990.

317. **Bennett, C. F., Balcarek, J. M., Varrichio, A., and Crooke, S. T.,** Molecular cloning and complete amino-acid sequence of form-I phosphoinositide-specific phospholipase C, *Nature,* 334, 268, 1988.

318. **Tsibris, J. C. M., Hunt, L. T., Ballejo, G., Barker, W. C., Toney, L. J., and Spellacy, W. N.,** Selective inhibition of protein disulfide isomerase by estrogens, *J. Biol. Chem.*, 264, 13967, 1989.

319. **Roth, M. A. and Pierce, S. B.,** *In vivo* crosslinking of protein disulphide isomerase to immunoglobulins, *Biochemistry,* 26, 4179, 1987.

320. **Lilie, H., McLaughlin, S., Freedman, R., and Buchner, J.,** Influence of protein disulfide isomerase (PDI) on antibody folding *in vitro*, *J. Biol. Chem.*, 269, 14290, 1994.

321. **Shorrosh, B. S., Subramaniam, J., Schubert, K. R., and Dixon, R. A.,** Expression and localization of plant protein disulfide isomerase, *Plant Physiol.*, 103, 719, 1993.

322. **Lambert, N. and Freedman, R. B.,** Structural properties of homogeneous protein disulphide isomerase from bovine liver purified by a rapid high-yielding procedure, *Biochem. J.*, 213, 225, 1983.

323. **Livesly, M. A., Bulleid, N. J., and Bray, C. M.,** Protein disulfide isomerase in germinating wheat (*Triticum aestivum*) seed during loss of viability, *Seed Sci. Res.*, 2, 97, 1992.

324. **Noiva, R. and Lennarz, W. J.,** Protein disulfide isomerase. A multi-functional protein resident in the lumen of the endoplasmic reticulum, *J. Biol. Chem.*, 267, 3553, 1992.

325. **Chrispeels, M. J.,** Biosynthesis of cell wall protein: sequential hydroxylation of proline, glycosylation of hydroxyproline and secretion of the glycoprotein, *Biochem. Biophys. Res. Commun.*, 39, 732, 1970.

326. **Andreae, M., Blankenstein, P., Zhang, Y.-H., and Robinson, D. G.,** Towards the subcellular localization of plant prolyl hydroxylase, *Eur. J. Cell Biol.*, 47, 181, 1988.

327. **Kieliszewski, M. and Lamport, D. T. A.,** Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny, *Plant J.*, 5, 157, 1994.

328. **Schafer, W. R. and Rine, J.,** Protein prenylation: genes, enzymes, targets, and functions, *Annu. Rev. Genet.*, 26, 209, 1992.

329. **Randall, S. K., Marshall, M. S., and Crowell, D. N.,** Protein isoprenylation in suspension-cultured tobacco cells, *Plant Cell,* 5, 433, 1993.

330. **Chrispeels, M. J. and Tague, B. W.,** Transport and targeting of proteins to protein storage vacuoles (protein bodies) in developing seeds, in *Recent Advances in Development and Germination of Seeds,* Taylorson, R. B., Ed., Plenum, New York, 1990.

331. **Staswick, P. E., Hermodson, M. A., and Nielsen, N. C.,** Identification of the cystines which link the acidic and basic components of the glycinin subunits, *J. Biol. Chem.*, 259, 13431, 1984.

332. **Krochko, J. E. and Bewley, J. D.,** Use of electrophoretic techniques in determining the composition of seed storage proteins in alfalfa, *Electrophoresis,* 9, 751, 1988.

333. **Dickinson, C. D., Scott, P. M., Hussein, E. H. A., Argos, P., and Nielsen, N. C.,** Effect of structural modifications on the assembly of a glycinin subunit, *Plant Cell,* 2, 403, 1990.

334. **Scott, M. P., Jung, R., Muntz, K., and Nielsen, N. C.,** A protease responsible for post-translational cleavage of a conserved Asn-Gly linkage in glycinin, the major seed storage protein of soybean, *Proc. Natl. Acad. Sci. U.S.A.*, 89, 658, 1992.

335. **Muramatsu, M. and Fukazawa, C.,** A high-order structure of plant storage proprotein allows its second conversion by an asparagine-specific cysteine protease, a novel proteolytic enzyme, *Eur. J. Biochem.*, 215, 123, 1993.

336. **Shimada, T., Hiraiwa, N., Nishimura, M., and Hara-Nishimura, I.,** Vacuolar processing enzyme that converts proproteins to the corresponding mature forms, *Plant Cell Physiol.*, 35, 713, 1994.

337. **D'Hondt, K., Bosch, D., Van Damme, J., Goethals, M., Vandekerckhove, J., and Krebbers, E.,** An aspartic proteinase present in seeds cleaves *Arabidopsis* 2S albumin precursors *in vitro*, *J. Biol. Chem.*, 268, 20884, 1993.

338. **Holwerda, B. C., Galvin, N. J., Baranski, T. J., and Rogers, J. C.,** *In vitro* processing of aleurain, a barley vacuolar thiol protease, *Plant Cell,* 2, 1091, 1990.

339. **Youle, R. and Huang, A. H. C.,** Occurrence of low molecular weight and high cysteine containing albumin storage protein in oil-seeds of diverse species, *Amer. J. Bot.*, 68, 44, 1981.

340. **Crouch, M. L., Tenbarge, K. M., Simon, A. E., and Ferl, R.,** cDNA clones for *Brassica napus* seed proteins: evidence from nucleotide sequence analysis that both subunits of napin are cleaved from a precursor polypeptide, *J. Mol. Appl. Gen.*, 2, 273, 1983.

341. **Sun, S. S. M., Altenbach, S. B., and Leung, F. W.,** Properties, biosynthesis and processing of a sulfur-rich protein in Brazil nut (*Bertholletia excelsa* H.B.K.), *Eur. J. Biochem.*, 162, 477, 1987.

342. **Hara-Nishimura, I., Takeuchi, Y., Inoue, K., and Nishimura, M.,** Vesicle transport and processing of the precursor to 2S albumin in pumpkin, *Plant J.*, 4, 793, 1993.

343. **Pueyo, J. J., Hunt, D. C., and Chrispeels, M. J.,** Activation of bean (*Phaseolus vulgaris*) α-amylase inhibitor requires proteolytic processing of the proprotein, *Plant Physiol.*, 101, 1341, 1993.

344. **Tanaka, T., Minamikawa, T., Yamauchi, D., and Ogushi, Y.,** Expression of an endopeptidase (EP-C1) in *Phaseolus vulgaris* plants, *Plant Physiol.*, 101, 421, 1993.

345. **Yamauchi, D., Akasofu, H., and Minamikawa, T.,** Cysteine endopeptidase from *Vigna mungo*: gene structure and expression, *Plant Cell Physiol.*, 33, 789, 1992.

346. **Qi, X., Chen, R., Wilson, K. A., and Tan-Wilson, A. L.,** Characterization of a soybean β-conglycinin-degrading protease cleavage site, *Plant Physiol.*, 104, 127, 1994.

347. **Qi, X., Wilson, K. A., and Tan-Wilson, A. L.,** Characterization of the major protease involved in the soybean β-conglycinin storage protein mobilization, *Plant Physiol.*, 99, 725, 1992.

348. **Mitsuhashi, W. and Oaks, A.,** Development of endopeptidase activities in maize (*Zea mays* L.) endosperms, *Plant Physiol.*, 104, 401, 1994.

349. **Abe, M., Arai, S., and Fujimaki, M.,** Purification and characterization of a protease occurring in endosperm of germinating corn, *Agric. Biol. Chem.*, 41, 893, 1977.

350. **De Barros, E. G. and Larkins, B. A.,** Purification and characterization of zein-degrading proteases from endosperm of germinating maize seeds, *Plant Physiol.*, 94, 297, 1990.

351. **De Barros, E. G. and Larkins, B. A.,** Cloning of a cDNA encoding a putative cysteine protease from germinating maize (*Zea mays* L.) seeds, *Intl. Soc. Plant Mol. Biol.*, Third Intl. Congress, Abstract 743, 1991.

352. **Yu, W.-J. and Greenwood, J. S.,** Purification and characterization of a cysteine proteinase from cotyledons of germinated *Vicia faba* L., *Plant Physiol.*, 102 (Suppl.), 53, 1993.

353. **Rothman, J. E.,** The compartmental organization of the Golgi appartus, *Sci. Am.*, 253(no. 3), 74, 1985.

354. **Griffiths, G. and Simons, K.,** The *trans*-Golgi network: sorting at the exit site of the Golgi complex, *Science,* 234, 438, 1986.

355. **Tooze, S. A. and Huttner, W. B.,** Cell-free protein sorting to the regulated and constitutive secretory pathways, *Cell,* 60, 837, 1990.

356. **Crowther, R. A., Finch, J. T., and Pearse, B. M. F.,** On the structure of coated vesicles, *J. Mol. Biol.*, 103, 785, 1976.

357. **Robinson, M. S.,** 100 kDa coated vesicle proteins: molecular heterogeneity and intracellular distribution studied with monoclonal antibodies, *J. Cell Biol.*, 204, 887, 1987.

358. **Robinson, D. G. and Depta, H.,** Coated vesicles, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 39, 53, 1988.

359. **Brodsky, F. M.,** Living with clathrin: its role in intracellular membrane traffic, *Science,* 242, 1396, 1988.

360. **Pearse, B. M. F.,** Clathrin and coated vesicles, *EMBO J.*, 6, 2507, 1987.

361. **Lobel, P., Fujimoto, K., Ye, R. D., Griffiths, G., and Kornfeld, S.,** Mutations in the cytoplasmic domain of the 275 kDa mannose 6-phosphate receptor differentially alter lysosome enzyme sorting and endocytosis, *Cell,* 57, 787, 1989.

362. **Glickman, J. N., Conibear, E., and Pearse, B. M. F.,** Specificity of binding of clathrin adaptors to signals on the mannose-6-phosphate insulin-like growth factor II receptor, *EMBO J.*, 8, 1041, 1989.

363. **Pearse, B. M. F.,** Receptors compete for adaptors found in plasma membrane coated pits, *EMBO J.*, 7, 3331, 1988.

364. **Pryer, N. K., Wuestehube, L. J., and Schekman, R.,** Vesicle-mediated protein sorting, *Annu. Rev. Biochem.*, 61, 471, 1992.

365. **Pearse, B. M. F. and Crowther, R. A.,** Structure and assembly of coated vesicles, *Annu. Rev. Biophys. Biophys. Chem.*, 16, 49, 1987.

366. **Brodsky, F. M., Hill, B. L., Acton, S. L., Nathke, I., and Wong, D.,** Clathrin light chains: arrays of protein motifs that regulate coated-vesicle dynamics, *Trends Biochem. Sci.*, 16, 208, 1991.

367. **Lin, H. C., Moore, M. S., Sanan, D. A., and Anderson, R. G. W.,** Reconstitution of clathrin-coated pit budding from plasma membranes, *J. Cell Biol.*, 114, 881, 1991.

368. **DeLuca-Flaherty, C., McKay, D. B., Parham, P., and Hill, B. L.,** Uncoating protein (hsc 70) binds a conformationally labile domain of clathrin light chain Lca to stimulate ATP hydrolysis, *Cell,* 62, 875, 1990.

369. **Rothman, J. E. and Schmid, S. L.,** Enzymatic recycling of clathrin from coated vesicles, *Cell,* 46, 5, 1986.

370. **Pearse, B. M. F. and Robinson, M. S.,** Clathrin, adaptors, and sorting, *Annu. Rev. Cell Biol.*, 6, 151, 1990.

371. **Keen, J. H.,** Clathrin and associated assembly and disassembly proteins, *Annu. Rev. Biochem.*, 59, 415, 1990.

372. **Ahle, S., Mann, A., Eichelsbacher, U., and Ungewickell, E.,** Structural relationships between clathrin assembly proteins from the Golgi and the plasma membrane, *EMBO J.*, 7, 919, 1988.

373. **Vigers, G. P. A., Crowther, R. A., and Pearse, B. M. F.,** *EMBO J.*, 5, 2079, 1986.

374. **Hopkins, C. R.,** Selective membrane protein trafficking: vectorial flow and filter, *Trends Biochem. Sci.*, 17, 27, 1992.

375. **Orci, L., Halban, P., Amherdt, M., Ravazzola, M., Vassalli, J.-D., and Perrelet, A.,** A clathrin-coated, Golgi-related compartment of the insulin secreting cell accumulates proinsulin in the presence of monensin, *Cell,* 39, 39, 1984.

376. **Orci, L., Glick, B. S., and Rothman, J. E.,** A new type of coated vesicular carrier that appears not to contain clathrin: its possible role in protein transport within the Golgi stack, *Cell,* 46, 171, 1986.

377. **Griffiths, G., Pfeiffer, S., Simons, K., and Matlin, K.,** Exit of newly synthesized membrane proteins from the *trans* cisterna of the Golgi complex to the plasma membrane, *J. Cell Biol.*, 101, 949, 1985.

378. Hillmer, S., Freundt, H., and Robinson, D. G., The partially coated reticulum and its relationship to the Golgi apparatus in higher plant cells, *Eur. J. Cell Biol.*, 47, 206, 1988.
379. Coleman, J., Evans, D., and Hawes, C., Plant coated vesicles, *Plant Cell Environ.*, 11, 669, 1988.
380. Robinson, D. G., Balusek, K., and Freundt, H., Legumin antibodies recognize polypeptides in coated vesicles isolated from developing pea cotyledons, *Protoplasma*, 150, 79, 1989.
381. Harley, S. M. and Beevers, L., Coated vesicles are involved in the transport of storage proteins during seed development in *Pisum sativum* L., *Plant Physiol.*, 91, 674, 1989.
382. Hoh, B., Schauermann, G., and Robinson, D. G., Storage protein polypeptides in clathrin coated vesicle fractions from developing pea cotyledons are not due to endomembrane contamination, *J. Plant Physiol.*, 138, 309, 1991.
383. Hoh, B. and Robinson, D. G., The prominent 28-kDa polypeptide in clathrin coated vesicle fractions from developing pea cotyledons is contaminating ferritin, *Cell Biol. Int.*, 17, 551, 1993.
384. Coleman, J. O. D., Evans, D. E., Horsley, D., and Hawes, C. R., The molecular structure of plant clathrin and coated vesicles, in *Endocytosis, Exocytosis and Vesicle Traffic in Plants*, Hawes, C. R., Coleman, J. O. D., and Evans, D. E., Eds., SEB Seminar Series Vol. 45, Cambridge University Press, 1991, 41.
385. Lin, H. B., Harley, S. M., Butler, J. M., and Beevers, L., Multiplicity of clathrin light-chain-like polypeptides from developing pea (*Pisum sativum* L.) cotyledons, *J. Cell Sci.*, 103, 1127, 1992.
386. Demmer, A., Holstein, S. E. H., Hinz, G., Schauermann, G., and Robinson, D. G., Improved coated vesicle isolation allows better characterization of clathrin polypeptides, *J. Exp. Bot.*, 44, 23, 1993.
387. Holstein et al., unpublished; cited in Reference 383.
388. Kirsch, T. and Beevers, L., Uncoating of clathrin-coated vesicles by uncoating ATPase from developing peas, *Plant Physiol.*, 103, 205, 1993.
389. Barinaga, M., Secrets of secretion revealed, *Science*, 260, 487, 1993.
390. Kahn, R. A., Yucel, J. K., and Malhotra, V., ARF signaling: a potential role for phospholipase D in membrane traffic, *Cell*, 75, 1045, 1993.
391. Novick, P. and Brennwald, P., Friends and family: the role of the Rab GTPases in vesicular traffic, *Cell*, 75, 597, 1993.
392. Hurtley, S. M., Membrane proteins involved in targeted membrane fusion, *Trends Biochem. Sci.*, 18, 453, 1993.
393. Rothman, J. E. and Orci, L. E., Molecular dissection of the secretory pathway, *Nature*, 355, 409, 1992.
394. Takizawa, P. A. and Malhotra, V., Coatomers and SNAREs in promoting membrane traffic, *Cell*, 75, 593, 1993.
395. Brose, N., Membrane fusion takes excitatory turn: syntaxin, vesicle docking protein, or glutamate receptor?, *Cell*, 75, 1043, 1993.
396. Burgoyne, R. D., Phosphoinositides in vesicular traffic, *Trends Biochem. Sci.*, 19, 55, 1994.
397. Beckers, C. J. M., Keller, D. S., and Balch, W. E., Semi-intact cells permeable to macromolecules: use in reconstitution of protein transport from the endoplasmic reticulum to the Golgi complex, *Cell*, 50, 523, 1987.
398. Simons, K. and Virta, H., Perforated MDCK cells support intracellular transport, *EMBO J.*, 6, 2241, 1987.
399. Baker, D., Hicke, L., Rexach, M., Schleyer, M., and Schekman, R., Reconstitution of *sec* gene product-dependent intercompartmental protein transport, *Cell*, 54, 335, 1988.
400. Deshaies, R. and Schekman, R., A yeast mutant defective at an early stage in import of secretory protein precursors into the endoplasmic reticulum, *J. Cell Biol.*, 105, 633, 1987.
401. Ferro-Novick, S., Hansen, W., Schauer, I., and Schekman, R., Genes required for completion of import of proteins into the endoplasmic reticulum in yeast, *J. Cell Biol.*, 98, 44, 1984.
402. Novick, P., Field, C., and Schekman, R., Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway, *Cell*, 21, 205, 1980.
403. Schekman, R., Protein localization and membrane traffic in yeast, *Annu. Rev. Cell Biol.*, 1, 115, 1985.
404. Burgoyne, R. D., Yeast mutants illuminate the secretory pathway, *Trends Biochem. Sci.*, 13, 241, 1988.
405. Fries, E. and Rothman, J. E., Transport of vesicular stomatitis virus glycoprotein in a cell-free extract, *Proc. Natl. Acad. Sci. U.S.A.*, 77, 3870, 1980.
406. Balch, W. E., Dunphy, W. G., Braell, W. A., and Rothman, J. E., Reconstitution of the transport of protein between successive compartments of the Golgi measured by the coupled incorporation of N-acetylglucosamine, *Cell*, 39, 405, 1984.
407. Balch, W. E. and Rothman, J. E., Characterization of protein transport between successive compartments of the Golgi apparatus: asymmetric properties of donor and acceptor activities in a cell-free system, *Arch. Biochem. Biophys.*, 240, 413, 1985.
408. Dunphy, W. G., Pfeffer, S. R., Clary, D. O., Wattenberg, B. W., Glick, B. S., and Rothman, J. E., Yeast and mammals utilize similar cytosolic components to drive protein transport through the Golgi complex, *Proc. Natl. Acad. Sci. U.S.A.*, 83, 1622, 1986.
409. Pâquet, M. R., Pfeffer, S. R., Burczak, J. D., Glick, B. S., and Rothman, J. E., Components responsible for transport between successive Golgi cisternae are highly conserved in evolution, *J. Biol. Chem.*, 261, 4367, 1986.
410. Hellgren, L., Morré, D. J., Selldén, G., and Sandelius, A. S., Isolation of a putative vesicular intermediate in the cell-free transfer of membrane from transitional endoplasmic reticulum to the Golgi apparatus of etiolated seedlings of garden pea, *J. Exp. Bot.*, 44 (Suppl.), 197, 1993.
411. Malhotra, V., Orci, L., Glick, B. S., Block, M. R., and Rothman, J. E., Role of an N-ethylmaleimide-sensitive transport component in promoting fusion of transport vesicles with cisternae of the Golgi stack, *Cell*, 54, 221, 1988.
412. Schekman, R., Genetic and biochemical analysis of vesicular traffic in yeast, *Curr. Opin. Cell Biol.*, 4, 587, 1992.
413. Bowser, R., Müller, H., Govindan, B., and Novick, P., Sec8p and Sec15p are components of a plasma membrane-associated 19.5S particle that may function downstream of Sec4p to control exocytosis, *J. Cell Biol.*, 118, 1041, 1992.
414. Kahn, R. A., Kern, F. G., Clark, J., Gelmann, E. P., and Rulka, C., Human ADP-ribosylation factors, *J. Biol. Chem.*, 266, 2606, 1991.
415. Palmer, D. J., Helms, J. B., Beckers, C. J. M., Orci, L., and Rothman, J. E., Binding of coatomer to Golgi membranes requires ADP-ribosylation factor, *J. Biol. Chem.*, 268, 12083, 1993.
416. Söllner, T., Whiteheart, S. W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., and Rothman, J. E., SNAP receptors implicated in vesicle targeting and fusion, *Nature*, 362, 318, 1993.
417. Verma, D. P. S., Cheon, C.-I., and Hong, Z., Small GTP-binding proteins and membrane biogenesis in plants, *Plant Physiol.*, 106, 1, 1994.
418. Südhof, T. C., De Camilli, P., Niemann, H., and Jahn, R., Membrane fusion machinery: insights from synaptic proteins, *Cell*, 75, 1, 1993.
419. Bennett, M. K., Garcia-Arraras, J. E., Elferink, L. A., Peterson, K., Fleming, A. M., Hazuka, C. D., and Scheller, R. H., The syntaxin family of vesicular transport receptors, *Cell*, 74, 863, 1993.
420. Brose, N., Petrenko, A. G., Südhof, T. C., and Jahn, R., Synaptotagmin: a calcium sensor on the synaptic vesicle surface, *Science*, 256, 1021, 1992.
421. Lian, J. P. and Ferro-Novick, S., Bos1p, an integral membrane protein of the endoplasmic reticulum to Golgi transport vesicles, is required for their fusion competence, *Cell*, 73, 735, 1993.
422. Hardwick, K. G. and Pelham, H. R. B., SED5 encodes a 39-kDa integral membrane protein required for vesicular transport between the ER and Golgi complex, *J. Cell Biol.*, 119, 513, 1992.
423. Burgoyne, R. D., Trimeric G proteins in Golgi transport, *Trends Biochem. Sci.*, 17, 87, 1992.
424. Hurtley, S. M., Hot line to the secretory pathway, *Trends Biochem. Sci.*, 18, 3, 1993.
425. Pfeffer, S. R., GTP-binding proteins in intracellular transport, *Trends Cell Biol.*, 2, 41, 1992.
426. Ferro-Novick, S. and Novick, P., The role of GTP-binding proteins in transport along the exocytic pathway, *Annu. Rev. Cell Biol.*, 9, 575, 1993.
427. Bourne, H. R., Do GTPases direct membrane traffic in secretion?, *Cell*, 53, 669, 1988.
428. Melancon, P., Glick, B. S., Malhotra, V., Weidman, P. J., Serafini, T., Gleason, M. L., Orci, L., and Rothman, J. E., Involvement of GTP-binding "G" proteins in transport through the Golgi stack, *Cell*, 51, 1053, 1987.
429. Bourne, H. R., Sanders, D. A., and McCormick, F., The GTPase super-family: a conserved switch for diverse cell functions, *Nature*, 348, 125, 1990.
430. Taylor, T. C., Kahn, R. A., and Melancon, P., Two distinct members of the ADP-ribosylation factor family of GTP-binding proteins regulate cell-free intra-Golgi transport, *Cell*, 70, 69, 1992.
431. Barlowe, C. and Schekman, R., *Sec12* encodes a guanine-nucleotide-exchange factor essential for transport vesicle budding from the ER, *Nature*, 365, 347, 1993.
432. Yoshihisa, Y., Barlowe, C., and Schekman, R., Requirement for a GTPase-activating protein in vesicle budding from the endoplasmic reticulum, *Science*, 259, 1466, 1993.
433. Kibbe, W. A., Hengst, L., and Gallwitz, D., in *The Ras Superfamily of GTPases*, Laca, J. C. and McCormick, F., Eds., CRC Press, Boca Raton, FL, 1993, 367.
434. Haubruck, H., Disela, C., Wagner, P., and Gallwitz, D., The *ras*-related *ypt* protein is a ubiquitous eukaryotic protein: isolation and sequence analysis of mouse cDNA clones highly homologous to the yeast YPT1 gene, *EMBO J.*, 6, 4049, 1987.
435. Segev, N., Mulholland, J., and Botstein, D., The yeast GTP-binding YPT1 protein and a mammalian counterpart are associated with the secretion machinery, *Cell*, 52, 915, 1988.
436. Touchot, N., Chardin, P., and Tavitian, A., Four additional members of the *ras* gene superfamily isolated by an oligonucleotide strategy: molecular cloning of YPT-related cDNAs from a rat brain library, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 8210, 1987.
437. Tisdale, E. J., Bourne, J. R., Khosravi-Far, R., Der, C. J., and Balch, W. E., GTP-binding mutants of Rab1 and Rab2 are potent inhibitors of vesicular transport from the endoplasmic reticulum to the Golgi complex, *J. Cell Biol.*, 119, 749, 1992.
438. Davidson, H. W. and Balch, W. E., Differential inhibition of multiple vesicular transport steps between the endoplasmic reticulum and *trans*-Golgi network, *J. Biol. Chem.*, 265, 4216, 1993.
439. Goud, B., Salminen, A., Walworth, N. C., and Novick, P. J., A GTP-binding protein required for secretion rapidly associates with secretory vesicles and the plasma membrane in yeast, *Cell*, 53, 753, 1988.
440. Borg, S. and Poulsen, C., Molecular analysis of two ypt/Rab-related sequences isolated from soybean (*Glycine max*) DNA libraries, *Plant Mol. Biol.*, 26, 175, 1994.

441. **Nagano, Y., Murai, N., Matsuno, R., and Sasaki, Y.**, Isolation and characterization of cDNAs that encode eleven small GTP-binding proteins from *Pisum sativum*, *Plant Cell Physiol.*, 34, 447, 1993.
442. **Palme, K., Diefenthal, T., Vingron, M., Sander, C., and Shell, J.**, Molecular cloning and structural analysis of genes from *Zea mays* (L.) coding for members of the *ras*-related *ypt* gene family, *Proc. Natl. Acad. Sci. U.S.A.*, 89, 787, 1992.
443. **Palme, K., Diefenthal, T., and Moore, I.**, The *ypt* gene family from maize and *Arapidopsis*: structural and functional analysis, *J. Exp. Bot.*, 44 (Suppl.), 183, 1993.
444. **Terryn, N., Van Montagu, M., and Inze, D.**, GTP-binding proteins in plants, *Plant Mol. Biol.*, 22, 143, 1993.
445. **Youssefian, S., Nakamura, M., and Sano, H.**, Molecular characterization of *rgp2*, a gene encoding a small GTP-binding protein from rice, *Mol. Gen. Genet.*, 237, 187, 1993.
446. **Bednarek, S. Y., Reynolds, T. L., Schroeder, M., Grabowski, R., Hengst, L., Gallwitz, D., and Raikhel, N. V.**, A small GTP-binding protein from *Arabidopsis thaliana* functionally complements the yeast *YPT6* null mutant, *Plant Physiol.*, 104, 591, 1994.
447. **Lebas, M. and Axelos, M.**, A cDNA encoding a new GTP binding protein of the ADP-ribosylation factor family from *Arabidopsis*, *Plant Physiol.*, 106, 809, 1994.
448. **Regad, F., Bardet, C., Tremousaygue, D., Moisan, A., Lescure, B., and Axelos, M.**, cDNA cloning and expression of an *Arabidopsis* GTP-binding protein of the ARF family, *FEBS Lett.*, 316, 133, 1993.
449. **Davies, C.**, Cloning and characterization of a tomato GTPase-like gene related to yeast and *Arabidopsis* genes involved in vesicular transport, *Plant Mol. Biol.*, 24, 525, 1994.
450. **d'Enfert, C., Gensse, M., and Gaillardin, C.**, Fission yeast and a plant have functional homologues of the Sar 1 and Sec 12 proteins involved in ER to Golgi traffic in budding yeast, *EMBO J.*, 11, 4205, 1992.
451. **Goud, B., Zahraoui, A., Tavitian, A., and Saraste, J.**, Small GTP-binding protein associated with Golgi cisternae, *Nature*, 345, 553, 1990.
452. **Antony, C., Cibert, C., Geraud, G., Maria, A. S., Maro, B., Mayau, V., and Goud, B.**, A small GTP-binding protein rab6p is distributed from medial Golgi to *trans*-Golgi network as determined by a confocal microscopic approach, *J. Cell Sci.*, 103, 785, 1992.
453. **Hengst, L. and Gallwitz, D.**, unpublished; cited in Reference 446.
454. **Shimmen, T. and Tazawa, M.**, Control of cytoplasmic streaming by ATP, Mg^{2+} and cytochalasin B in permeabilised Characeae cell, *Protoplasma*, 115, 18, 1983.
455. **Tester, M., Beilby, M. J., and Shimmen, T.**, Electrical characteristics of the tonoplast of *Chara corallina*: a study using permeabilised cells, *Plant Cell Physiol.*, 28, 1555, 1987.
456. **Zhang, D. H., Wadsworth, P., and Hepler, P. K.**, Modulation of anaphase spindle microtubule structure in stamen hair cells of *Tradescantia* by calcium and related agents, *J. Cell Sci.*, 102, 79, 1992.
457. **Hillmer, S., Gilroy, S., and Jones, R. L.**, Visualizing enzyme secretion from individual barley (*Hordeum vulgare*) protoplasts, *Plant Physiol.*, 102, 279, 1992.
458. **Kamada, I., Yamauchi, S., Youssefian, S., and Sano, H.**, Transgenic tobacco plants expressing rgp1, a gene encoding a ras-related GTP-binding protein from rice, show distinct morphological characteristics, *Plant J.*, 2, 799, 1992.
459. **Barbacid, M.**, Ras genes, *Annu. Rev. Biochem.*, 56, 779, 1987.
460. **Bourne, H. R., Sanders, D. A., and McCormick, F.**, The GTPase superfamily: conserved structure and molecular mechanism, *Nature*, 349, 117, 1991.
461. **Takai, Y., Kaibuchi, K., Kikuchi, A., and Kawata, M.**, Small GTP-binding proteins, *Int. Rev. Cytol.*, 133, 187, 1992.
462. **Nuoffer, C., Davidson, H. W., Matteson, J., Meinkoth, J., and Balch, W. E.**, A GDP-bound form of Rab1 inhibits protein export from the endoplasmic reticulum and transport between Golgi compartments, *J. Cell Biol.*, 125, 225, 1994.
463. **Strom, M., Vollmer, P., Tan, T. J., and Gallwitz, D.**, A yeast GTPase-activating protein that interacts specifically with a member of the Ypt/Rab family, *Nature*, 361, 736, 1993.
464. **Moya, M., Roberts, D., and Novick, P.**, *Dss4-1* is a dominant suppressor of *sec4-8* that encodes a nucleotide exchange protein that aids Sec4p function, *Nature*, 361, 460, 1993.
465. **Pind, S. N., Nuoffer, C., McCaffery, J. M., Plutner, H., Davidson, H. W., Farquhar, M. G., and Balch, W. E.**, Rab1 and Ca^{2+} are required for the fusion of carrier vesicles mediating endoplasmic reticulum to Golgi transport, *J. Cell Biol.*, 125, 239, 1994.
466. **Zahner, J. E. and Cheney, C. M.**, A *Drosophila* homolog of bovine *smg* p25a GDP dissociation inhibitor undergoes a shift in isoelectric point in the developmental mutant quartet, *Mol. Cell. Biol.*, 13, 217, 1993.
467. **Garrett, M. D., Kabcenell, A. K., Zahner, J. E., Kaibuchi, K., Sasaki, T., Takai, Y., Cheney, C. M., and Novick, P. J.**, Interaction of Sec 4 with GDI proteins from bovine brain, *Drosophila melanogaster* and *Saccharomyces cerevisiae*, *FEBS Lett.*, 331, 233, 1993.
468. **Ullrich, O., Horiuchi, H., Bucci, C., and Zerial, M.**, Membrane association of Rab5 mediated by GDP-dissociation inhibitor and accompanied by GDP/GTP exchange, *Nature*, 368, 157, 1994.
469. **Chrispeels, M. J. and Greenwood, J. S.**, Heat stress enhances phytohemagglutinin synthesis but inhibits its transport out of the endoplasmic reticulum, *Plant Physiol.*, 83, 778, 1987.
470. **Altschuler, M. and Mascarenhas, J. P.**, Transcription and translation of heat shock and normal proteins in seedlings and developing seeds of soybean exposed to a gradual temperature increase, *Plant Mol. Biol.*, 5, 291, 1985.
471. **Bush, D. S., Sticher, L., Huystee, R., Wagner, D., and Jones, R. L.**, The calcium requirement for stability and enzymatic activity of two isoforms of barley aleurone α-amylase, *J. Biol. Chem.*, 264, 19392, 1989.
472. **Bush, D. S., Biswas, A. K., and Jones, R. L.**, Gibberellic-acid-stimulated Ca^{2+} accumulation in endoplasmic reticulum of barley aleurone: Ca^{2+} transport and steady state levels, *Planta*, 178, 411, 1989.
473. **Jones, R. L., Bush, D. S., and Biswas, A. K.**, Gibberellic acid and calcium participate in the synthesis of active α-amylase molecules, in *Growth Regulators and Seeds*, Pinfield, N. J. and Black, M., Eds., British Plant Regulator Group Monograph Vol. 15, 1987, 15, 31.
474. **Skriver, K., Olsen, F. L., Rogers, J. C., and Mundy, J.**, *Cis*-acting DNA elements responsive to gibberellin and its antagonist abscisic acid, *Proc. Natl. Acad. Sci. U.S.A.*, 88, 7266, 1991.
475. **Jacobsen, J. V. and Close, T. J.**, Control of transient expression of chimaeric genes by gibberellic acid and abscisic acid in protoplasts prepared from mature barley aleurone layers, *Plant Mol. Biol.*, 16, 713, 1991.
476. **Gubler, F. and Jacobsen, J. V.**, Gibberellin-responsive elements in the promoter of a barley high-pI α-amylase gene, *Plant Cell*, 4, 1435, 1992.
477. **Lanahan, M. B., Ho, T.-H. D., Rogers, S. W., and Rogers, J. C.**, A gibberellin response complex in cereal α-amylase gene promoters, *Plant Cell*, 4, 203, 1992.
478. **Rogers, J. C., Lanahan, M. B., and Rogers, S. W.**, The *cis*-acting gibberellin response complex in high-pI-α amylase gene promoters. Requirement for a coupling element for high-level transcription, *Plant Physiol.*, 105, 151, 1994.
479. **Gilroy, S. and Jones, R. L.**, Gibberellic acid and abscisic acid coordinately regulate cytoplasmic calcium and secretory activity in barley aleurone protoplasts, *Proc. Natl. Acad. Sci. U.S.A.*, 89, 3591, 1992.
480. **Bush, D. S. and Jones, R. L.**, Hormonal control of Ca^{2+} transport in microsomal vesicles isolated from barley aleurone layers, in *Calcium in Plant Growth and Development*, Leonard, R. T. and Hepler, P. K., Eds., American Society of Plant Physiologists, Rockville, MD, 1990, 60.
481. **DuPont, F. M., Bush, D. S., Windle, J. J., and Jones, R. L.**, Interactions between calcium and proton transport in membrane vesicles from barley roots, *Plant Physiol.*, 94, 179, 1990.
482. **Gilroy, S. and Jones, R. L.**, Calmodulin stimulation of unidirectional calcium uptake by the endoplasmic reticulum of barley aleurone, *Planta*, 190, 289, 1993.
483. **Allan, A. C. and Trewavas, A. J.**, Abscisic acid and gibberellin perception: inside or out?, *Plant Physiol.*, 104, 1107, 1994.
484. **Gilroy, S. and Jones, R. L.**, Perception of gibberellin and abscisic acid at the external face of the plasma membrane of barley (*Hordeum vulgare* L.) aleurone protoplasts, *Plant Physiol.*, 104, 1185, 1994.
485. **Hooley, R., Beale, M. H., and Smith, S. J.**, Gibberellin perception at the plasma membrane of *Avena fatua* aleurone protoplasts, *Planta*, 183, 274, 1991.
486. **Caspers, et al.**, unpublished; cited in Reference 4.
487. **Chrispeels, M. J. and Tague, B. W.**, Protein sorting in the secretory system of plant cells, *Int. Rev. Cytol.*, 125, 1, 1991.
488. **Gal, S. and Raikhel, N. V.**, Protein sorting in the endomembrane system of plant cells, *Curr. Opin. Cell Biol.*, 5, 636, 1993.
489. **Galili, G., Altschuler, Y., and Levanony, A.**, Assembly and transport of seed storage proteins, *Trends Cell Biol.*, 3, 437, 1993.
490. **Chrispeels, M. J. and Raikhel, N. V.**, Short peptide domains target proteins to plant vacuoles, *Cell*, 68, 613, 1992.
491. **Vitale, A. and Chrispeels, M. J.**, Sorting of proteins to the vacuoles of plant cells, *BioEssays*, 14, 151, 1992.
492. **Nakamura, K. and Matsuoka, K.**, Protein targeting to the vacuole in plant cells, *Plant Physiol.*, 101, 1, 1993.
493. **Herman, E. M., Hankins, C. N., and Shannon, L. M.**, Bark and leaf lectins of *Sophora japonica* are sequestered in protein-storage vacuoles, *Plant Physiol.*, 86, 1027, 1988.
494. **Hakman, I.**, Embryology in Norway spruce (*Picea abies*). Immunochemical studies on transport of a seed storage protein, *Physiol. Plant.*, 88, 427, 1993.
495. **Collada, C., Allona, I., Aragoncillo, P., and Aragoncillo, C.**, Development of protein bodies in cotyledons of *Fagus sylvatica*, *Physiol. Plant.*, 89, 354, 1993.
496. **Muntz, K.**, Intracellular protein sorting and the formation of protein reserves in storage tissue cells of plant seeds, *Biochem. Physiol. Pflanzen*, 185, 315, 1989.
497. **Harris, N.**, Organization of the endomembrane system, *Annu. Rev. Plant Physiol.*, 37, 73, 1986.
498. **Lending, C. R. and Larkins, B. A.**, Changes in the zein composition of protein bodies during maize endosperm development, *Plant Cell*, 1, 1011, 1989.
499. **Boller, T. and Wiemken, A.**, Dynamics of vacuolar compartmentation, *Annu. Rev. Plant Physiol.*, 37, 137, 1986.
500. **Kim, W. T., Franceschi, V. R., Krishnan, H. B., and Okita, T. W.**, Formation of wheat protein bodies: involvement of the Golgi apparatus in gliadin transport, *Planta*, 176, 173, 1988.
501. **Larkins, B. A. and Hurkman, W. J.**, Synthesis and deposition of zein in protein bodies of maize endosperm, *Plant Physiol.*, 62, 256, 1978.

502. **Krishnan, H. B., Franceschi, V. R., and Okita, T. W.**, Immunochemical studies on the role of the Golgi complex in protein-body formation in rice seeds, *Planta*, 169, 471, 1986.

503. **Taylor, J.R.N., Schussler, L., and Liebonberg, N.V.D.W.**, Protein body formation in the starchy endosperm of developing *Sorghum bicolor* (L) moench seeds, *S. Afr. J. Bot.*, 51, 35, 1985.

504. **Li, X., Wu, Y., Zhang, D.-Z., Gillikin, J. W., Boston, R. S., Franceschi, V. R., and Okita, T. W.**, Rice prolamine protein body biogenesis: a BiP-mediated process, *Science*, 262, 1054, 1993.

505. **Li, X., Franceschi, V. R., and Okita, T. W.**, Segregation of storage protein mRNAs on the rough endoplasmic reticulum membranes of rice endosperm cells, *Cell*, 72, 869, 1993.

506. **Abe, S., You, W., and Davies, E.**, Protein bodies in corn endosperm are enclosed by and enmeshed in F-actin, *Protoplasma*, 165, 139, 1991.

507. **Singer, R. H., Langevin, G. L., and Lawrence, J. B.**, Ultrastructural visualization of cytoskeletal mRNAs and their associated proteins using double-label *in situ* hybridization, *J. Cell Biol.*, 108, 2343, 1989.

507a. **Hoffman, L. M., Donaldson, D. D., Bookland, R., Rashka, K., and Herman, E. M.**, Synthesis and protein deposition of maize 15-kDa zein in transgenic tobacco seeds, *EMBO J.*, 6, 3213, 1987.

507b. **Bagga, S., Adams, H., Kemp, J. D., and Sengupta-Gopalan, C.**, Accumulation of 15-kilodalton zein in novel protein bodies in transgenic tobacco, *Plant Physiol.*, 107, 13, 1995.

508. **Larkins, B. A., Pedersen, K., Handa, A. K., Hurkman, W. J., and Smith, L. D.**, Synthesis and processing of maize storage proteins in *Xenopus laevis* oocytes, *Proc. Natl. Acad. Sci. U.S.A.*, 76, 6448, 1979.

509. **Hurkman, W. J., Smith, L. D., Richter, J., and Larkins, B. A.**, Subcellular compartmentalization of maize storage proteins in *Xenopus* oocytes injected with zein messenger RNAs, *J. Cell Biol.*, 89, 292, 1981.

510. **Simon, R., Altschuler, Y., Rubin, R., and Galili, G.**, Two closely related wheat storage proteins follow a markedly different subcellular route in *Xenopus laevis* oocytes, *Plant Cell*, 2, 941, 1990.

511. **Kim, W. T. and Okita, T. W.**, Structure, expression, and heterogeneity of the rice seed prolamines, *Plant Physiol.*, 88, 649, 1988.

512. **Shewry, P. R. and Tatham, A. S.**, The prolamin storage proteins of cereal seeds: structure and evolution, *Biochem. J.*, 267, 1, 1990.

513. **Kreis, M., Shewry, P. R., Ford, B. G., Ford, J., and Miflin, B. J.**, Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley and rye, in *Oxford Survey of Plant Molecular and Cellular Biology*, Vol. 2, Miflin, B. J., Ed., Oxford University Press, New York, 1985, 253.

514. **Altschuler, Y., Rosenberg, N., Harel, R., and Galili, G.**, The N- and C-terminal regions regulate the transport of wheat γ-gliadin through the endoplasmic reticulum in *Xenopus* oocytes, *Plant Cell*, 5, 443, 1993.

515. **Torrent, M., Geli, M. I., Ruiz-Avila, L., Canals, J. M., Puigdomènech, P., and Ludevid, D.**, Role of structural domains for maize γ-zein retention in *Xenopus* oocytes, *Planta*, 192, 512, 1994.

516. **Rubin, R., Levanony, H., and Galili, G.**, Characterization of two types of protein bodies in developing wheat endosperm, *Plant Physiol.*, 99, 718, 1992.

517. **Levanony, H., Rubin, R., Altschuler, Y., and Galili, G.**, Evidence for a novel route of wheat storage proteins to vacuoles, *J. Cell Biol.*, 119, 1117, 1992.

518. **Vitale, A., Sturm, A., and Bollini, R.**, Regulation of processing of a plant glycoprotein in the Golgi complex: a comparative study using *Xenopus* oocytes, *Planta*, 169, 108, 1986.

519. **Voelker, T. A., Florkiewicz, R. Z., and Chrispeels, M. J.**, Secretion of phytohemagglutinin by monkey COS cells, *Eur. J. Cell Biol.*, 42, 218, 1986.

520. **Bustos, M. M., Luckow, V. A., Griffing, L. R., Summers, M. D., and Hall, T. C.**, Expression, glycosylation and secretion of phaseolin in a baculovirus system, *Plant Mol. Biol.*, 10, 475, 1988.

521. **Tague, B. W. and Chrispeels, M. J.**, The plant vacuolar protein, phytohemagglutinin, is transported to the vacuole of transgenic yeast, *J. Cell Biol.*, 105, 1971, 1987.

522. **Woolford, C. A., Daniels, L. B., Park, F. J., Jones, E. W., van Arsdell, J. N., and Innis, M. A.**, The PEP4 gene encodes an asparatyl protease implicated in the post-translational processing of *Saccharomyces cerevisiae* vacuolar hydrolases, *Mol. Cell. Biol.*, 6, 2500, 1986.

523. **Rothman, J. H., Yamashiro, C. T., Kane, P. M., and Stevens, T. H.**, Protein targeting to the yeast vacuole, *Trends Biochem. Sci.*, 14, 347, 1989.

524. **Klionsky, D. J., Herman, P. K., and Emr, S. D.**, The fungal vacuole: composition, function, and biogenesis, *Microbiol. Rev.*, 54, 266, 1990.

525. **Valls, L. A., Hunter, C. P., Rothman, J. H., and Stevens, T. H.**, Protein sorting in yeast: the localization determinant of yeast vacuolar carboxypeptidase Y resides in the propeptide, *Cell*, 48, 887, 1987.

526. **Valls, L. A.**, A Mutational Analysis of a Yeast Vacuolar Sorting Determinant, Ph.D. dissertation, University of Oregon, 1988.

527. **Johnson, L. M., Bankaitis, V. A., and Emr, S. D.**, Distinct sequence determinants direct intracellular sorting and modification of a yeast vacuolar protease, *Cell*, 48, 875, 1987.

528. **Valls, L. A., Winter, J. R., and Stevens, T. H.**, Yeast carboxypeptidase Y vacuolar targeting signal is defined by four propeptide amino acids, *J. Cell Biol.*, 111, 361, 1990.

529. **Klionsky, D. J., Banta, L. M., and Emr, S. D.**, Intracellular sorting and processing of a yeast vacuolar hydrolase: proteinase A propeptide contains vacuolar targeting information, *Mol. Cell. Biol.*, 8, 2105, 1988.

530. **Moehle, C. M., Dixon, C. K., and Jones, R. B.**, Processing pathway for protease B of *Saccharomyces cerevisiae*, J. *Cell Biol.*, 108, 309, 1989.

531. **Klionsky, D. J. and Emr, S. D.**, Membrane protein sorting: biosynthesis, transport, and processing of yeast vacuolar alkaline phosphatase, *EMBO J.*, 8, 2241, 1989.

532. **Roberts, C. J., Pohlig, G., Rothman, J. H., and Stevens, T. H.**, Structure, biosynthesis, and localization of dipeptidyl amino peptidase B, an integral membrane glycoprotein of the yeast vacuole, *J. Cell Biol.*, 108, 1363, 1989.

533. **Roberts, C. J., Nothwehr, S. F., and Stevens, T. H.**, Membrane protein sorting in the yeast secretory pathway. Evidence that the vacuole may be the default compartment, *J. Cell Biol.*, 119, 69, 1992.

534. **Yoshihisa, T. and Anraku, Y.**, A novel pathway of import of α-mannosidase, a marker enzyme of vacuolar membrane, in *Saccharomyces cerevisiae*, *J. Biol. Chem.*, 265, 22418, 1990.

535. **Bankaitis, V. A., Johnson, L. M., and Emr, S. D.**, Isolation of yeast mutants defective in protein targeting to the vacuole, *Proc. Natl. Acad. Sci. U.S.A.*, 83, 9075, 1986.

536. **Rothman, J. H. and Stevens, T. H.**, Protein sorting in yeast: mutants defective in vacuole biogenesis mislocalize vacuolar proteins into the late secretory pathway, *Cell*, 47, 1041, 1986.

537. **Robinson, J. S., Klionsky, D. J., Banta, L. M., and Emr, S. D.**, Protein sorting in *Saccharomyces cerevisiae*: isolation of mutants defective in the delivery and processing of multiple vacuolar hydrolases, *Mol. Cell. Biol.*, 8, 4936, 1988.

538. **Rothman, J. H., Raymond, C. K., Gilbert, T., O'Hara, P. J., and Stevens, T. H.**, A putative GTP binding protein homologous to interferon-inducible Mx proteins performs an essential function in yeast protein sorting, *Cell*, 61, 1063, 1990.

539. **Herman, P. K., Stack, J. H., DeModena, J. A., and Emr, S. D.**, A novel protein kinase homolog essential for protein sorting to the yeast lysosome-like vacuole, *Cell*, 64, 425, 1991.

540. **Herman, P. K., Stack, J. H., and Emr, S.**, A genetic and structural analysis of the yeast Vps 15 protein kinase: evidence for a direct role of Vps15p in vacuolar protein delivery, *EMBO J.*, 10, 4049, 1991.

541. **Felder, S., Miller, K., Moehren, G., Ullrich, A., Schlessinger, J., and Hopkins, C. R.**, Kinase activity controls the sorting of the epidermal growth factor receptor within the multivesicular body, *Cell*, 61, 623, 1990.

542. **Honegger, A. M., Schmidt, A., Ullrich, A., and Schlessinger, J.**, Separate endocytic pathways of kinase-defective and -active EGF receptor mutants expressed in the same cells, *J. Cell Biol.*, 110, 1541, 1990.

543. **Herman, P. K. and Emr, S. D.**, Characterization of VPS34, a gene required for vacuolar protein sorting and vacuolar segregation in *Saccharomyces cerevisiae*, *Mol. Cell. Biol.*, 10, 6742, 1990.

544. **Schu, P. V., Takegawa, K., Fry, M. J., Stack, J. H., Waterfield, M. D., and Emr, S. D.**, Phosphatidylinositol 3-kinase encoded by yeast *vps34* gene essential for protein sorting, *Science*, 260, 88, 1993.

545. **Welters, P., Takegawa, K., Emr, S. D., and Chrispeels, M. J.**, Isolation of an *Arabidopsis thaliana* gene homolog to a yeast gene needed for correct vacuolar targeting, *Plant Physiol.*, Suppl., 102, 149, 1993.

546. **Singer-Krüger, B., Stenmark, H., Düsterhöft, A., Philippsen, P., Yoo, J.-S., Gallwitz, D., and Zerial, M.**, Role of three Rab5-like GTPases, Ypt51p, Ypt52p, and Ypt53p, in the endocytic and vacuolar protein sorting pathways of yeast, *J. Cell Biol.*, 125, 283, 1994.

547. **Shimada, T., Nishimura, M., and Hara-Nishimura, I.**, Small GTP-binding proteins are associated with the vesicles that are targeted to vacuoles in developing pumpkin cotyledons, *Plant Cell Physiol.*, 35, 995, 1994.

548. **Horazdovsky, B. F. and Emr, S. D.**, The *vps16* gene product associates with a sedimentable protein complex and is essential for vacuolar protein sorting in yeast, *J. Biol. Chem.*, 268, 4953, 1993.

549. **Rothman, J. H., Howald, I., and Stevens, T. H.**, Characterization of genes required for protein sorting and vacuolar function in the yeast *Saccharomyces cerevisiae*, *EMBO J.*, 8, 2057, 1989.

550. **Banta, L. M., Robinson, J. S., Klionsky, D. J., and Emr, S. D.**, Organelle assembly in yeast: characterization of yeast mutants defective in vacuolar biogenesis and protein sorting, *J. Cell Biol.*, 107, 1369, 1988.

551. **Hill, K. J. and Stevens, T. H.**, Vma21p is a yeast membrane protein retained in the endoplasmic reticulum by a di-lysine motif and is required for the assembly of the vacuolar H⁺-ATPase complex, *Mol. Biol. Cell*, 5, 1039, 1994.

552. **Bachhawat, A. K., Manolson, M. F., Murdock, D. G., Garman, J. D., and Jones, E. W.**, The *vph2* gene encodes a 25 kDa protein required for activity of the yeast vacuolar H⁺-ATPase, *Yeast*, 9, 175, 1993.

553. **Ho, M. N., Hill, K. J., Lindorfer, M. A., and Stevens, T. H.**, Isolation of vacuolar membrane H⁺-ATPase-deficient mutants; the *VMA5* and *VMA4* genes are essential for assembly and activity of the vacuolar H⁺-ATPase, *J. Biol. Chem.*, 268, 221, 1993.

554. **Ho, M. N., Hirata, R., Umemoto, N., Ohya, Y., Takatsuki, A., Stevens, T. H., and Anraku, Y.**, *VMA13* encodes a 54 kDa vacuolar H⁺-ATPase subunit required for activity but not assembly of the enzyme complex, *J. Biol. Chem.*, 268, 18286, 1993.

555. **Forgac, M.,** Structure and function of vacuolar class of ATP driven proton pumps, *Physiol. Rev.*, 69, 765, 1989.
556. **Tague, B. W. and Goodman, H. M.,** Approaches to the isolation of pH mutants in *Arabidopsis*, *Intl. Soc. Plant Mol. Biol.*, Third Intl. Congress, Abstract 1014, 1991.
557. **Vida, T. A., Graham, T. R., and Emr, S. D.,** *In vitro* reconstitution of intercompartmental protein transport to the yeast vacuole, *J. Cell Biol.*, 111, 2871, 1990.
558. **Tague, B. W., Dickinson, C. D., and Chrispeels, M. J.,** A short domain of the plant vacuolar protein phytohemagglutinin targets invertase to the yeast vacuole, *Plant Cell*, 2, 533, 1990.
559. **Von Schaewen, A. and Chrispeels, M. J.,** Identification of vacuolar sorting information in phytohemagglutinin, an unprocessed vacuolar protein, *J. Exp. Bot.*, 44 (Suppl.), 339, 1993.
560. **Saalbach, G., Jung, R., Kunze, G., Saalbach, I., Adler, K., and Muntz, K.,** Different legumin protein domains act as vacuolar targeting signals, *Plant Cell*, 3, 695, 1991.
561. **Gal, S. and Raikhel, N. V.,** A carboxy-terminal plant vacuolar targeting signal is not recognized in yeast, *Plant J.*, 6, 235, 1994.
562. **Matsuoka, K. and Nakamura, K.,** Transport of a sweet potato storage protein, sporamin, to the vacuole in yeast cells, *Plant Cell Physiol.*, 33, 453, 1992.
563. **Bednarek, S. Y., Wilkins, T. A., Dombrowski, J. E., and Raikhel, N. V.,** A carboxyl-terminal propeptide is necessary for proper sorting of barley lectin to vacuoles of tobacco, *Plant Cell*, 2, 1145, 1990.
564. **Bednarek, S. Y. and Raikhel, N. V.,** The barley lectin carboxyl-terminal propeptide is a vacuolar sorting determinant in plants, *Plant Cell*, 3, 1195, 1991.
565. **Schroeder, M. R., Dombrowski, J. E., Bednarek, S. Y., Borkhsenious, O. N. and Raikhel, N. V.,** Molecular basis of post-translational modifications and targeting of barley lectin to the vacuoles in barley and in transgenic tobacco plants, *J. Exp. Bot.*, 44 (Suppl.), 315, 1993.
566. **Dombrowski, J. E., Schroeder, M. R., Bednarek, S. Y., and Raikhel, N. V.,** Determination of the functional elements within the vacuolar targeting signal of barley lectin, *Plant Cell*, 5, 587, 1993.
567. **Lerner, D. R. and Raikhel, N. V.,** Cloning and characterization of root-specific barley lectin, *Plant Physiol.*, 91, 124, 1989.
568. **Raikhel, N. V. and Lerner, D. R.,** Expression and regulation of lectin genes in cereals and rice, *Dev. Genet.*, 12, 255, 1991.
569. **Hattori, T., Ichihara, S., and Nakamura, K.,** Processing of a plant vacuolar protein precursor *in vitro*, *Eur. J. Biochem.*, 166, 533, 1987.
570. **Matsuoka, K. and Nakamura, K.,** Propeptide of a precursor to a plant vacuolar protein required for vacuolar targeting, *Proc. Natl. Acad. Sci. U.S.A.*, 88, 834, 1991.
571. **Wilkins, T. A. and Raikhel, N. V.,** Expression of rice lectin is governed by two temporally and spatially regulated mRNAs in developing embryos, *Plant Cell*, 1, 541, 1989.
572. **Nakamura, K., Matsuoka, K., Mukumoto, F., and Watanabe, N.,** Processing and transport to the vacuole of a precursor to sweet potato sporamin in transformed tobacco cell line BY-2, *J. Exp. Bot.*, 44 (Suppl.), 331, 1993.
573. **Matsuoka, K. and Nakamura, K.,** Personal communication.
574. **Saalbach, G., Schumann, U., Saalbach, I., and Müntz, K.,** Personal communication.
575. **D'Hondt, K., Van Damme, J., Van Den Bossche, C., Leejeerajumneun, S., De Rycke, K., Derksen, J., Vandekerckhove, J., and Krebbers, E.,** Studies of the role of the propeptides of the *Arabidopsis thaliana* 2S albumin, *Plant Physiol.*, 102, 425, 1993.
576. **Boller, T. and Vogell, U.,** Vacuolar localization of ethylene-induced chitinase in bean leaves, *Plant Physiol.*, 74, 442, 1984.
577. **Mauch, F. and Staehelin, L. A.,** Functional implications of the subcellular localization of ethylene-induced chitinase and β-1,3 glucanase in bean leaves, *Plant Cell*, 1, 447, 1989.
578. **Neuhaus, J.-M., Sticher, L., Meins, F., Jr., and Boller, T.,** A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole, *Proc. Natl. Acad. Sci. U.S.A.*, 88, 10362, 1991.
579. **Melchers, L. S., Sela-Buurlage, M. B., Vloemans, S. A., Woloshuk, C. P., Van Roekel, J. S. C., Pen, J., and van den Elzen, P. J. M.,** Extracellular targeting of the vacuolar tobacco proteins AP24, chitinase and β-1,3-glucanase in transgenic plants, *Plant Mol. Biol.*, 21, 583, 1993.
580. **Neuhaus, J.-M., Pietrzak, M., and Boller, T.,** Mutation analysis of the C-terminal vacuolar targeting peptide of tobacco chitinase: low specificity of the sorting system, and gradual transition between intracellular retention and secretion into the extracellular space, *Plant J.*, 5, 45, 1994.
581. **Holwerda, B. C. and Rogers, J. C.,** Structure, functional properties and vacuolar targeting of the barley thiol protease, aleurain, *J. Exp. Bot.*, 44 (Suppl), 321, 1993.
582. **Miao, G.-H., Hong, Z., and Verma, D. P. S.,** Topology and phosphorylation of soybean nodulin-26, an intrinsic protein of the peribacteroid membrane, *J. Cell Biol.*, 118, 481, 1992.
583. **Cheon, C.-I., Hong, Z., and Verma, D. P. S.,** Nodulin-24 follows a novel pathway for integration into the peribacteroid membrane in soybean root nodules, *J. Biol. Chem.*, 269, 6598, 1994.
584. **Kermode, A. R., Fisher, A. S., Polishchuk, E., Wandelt, C., Spencer, D., and Higgins, T. J. V.,** Accumulation and proteolytic processing of vicilin-deletion mutant proteins in the leaf and seed of transgenic tobacco, *Planta*, 197, 501, 1995.
585. **Nakamura, L, Dube, P. H., and Beachy, R. N.,** Accumulation of the products of β-conglycinin α′-subunit gene constitutively expressed in seeds and non-seed tissues of the transgenic petunia plants, *Plant Cell Physiol.*, 34, 865, 1993.
586. **Krochko, J. E., Bantroch, D. J., Greenwood, J. S., and Bewley, J. D.,** Seed storage proteins in developing somatic embryos of alfalfa: defects in accumulation compared to zygotic embryos, *J. Exp. Bot.*, 45, 699, 1994.
587. **Cooper, A. and Bussey, H.,** Yeast Kex1p is a Golgi-associated membrane protein: Deletions result in mislocalization to the vacuolar membrane, *J. Cell Biol.*, 119, 1459, 1992.
588. **Donaldson, J. G., Finazzi, D., and Klausner, R. D.,** Brefeldin A inhibits Golgi membrane-catalysed exchange of guanine nucleotide onto ARF protein, *Nature*, 360, 350, 1992.
589. **Helms, J. B. and Rothman, J. E.,** Inhibition by brefeldin A of a Golgi membrane enzyme that catalyses exchange of guanine nucleotide bound to ARF, *Nature*, 360, 352, 1992.
590. **Hinz, G., Hoh, B., and Robinson, D. G.,** Strategies in the recognition and isolation of storage protein receptors, *J. Exp. Bot.*, 44 (Suppl.), 351, 1993.
591. **Kirsch, T., Paris, N., Butler, J. M., Beevers, L., and Rogers, J. C.,** Purification and initial characterization of a potential plant vacuolar targeting receptor, *Proc. Natl. Acad. Sci. U.S.A.*, 91, 3403, 1994.
591a. **Okita, T. W. and Rogers, J. C.,** Compartmentation of proteins in the endomembrane system of plant cells, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, in press.
592. **Schroeder, R. M., Borkhsenious, O. N., Matsuoka, K., Nakamura, K., and Raikhel, N. V.,** Colocalization of barley lectin and sporamin in vacuoles of transgenic tobacco plants, *Plant Physiol.*, 101, 451, 1993.
593. **Eckes, P., Schmitt, P., Daub, W., and Wengenmayer, F.,** Overproduction of alfalfa glutamine synthetase in transgenic tobacco plants, *Mol. Gen. Genet.*, 217, 263, 1989.
594. **Hilder, V. A., Gatehouse, A. M. R., Sheerman, S. E., Barker, R. F., and Boulter, D.,** A novel mechanism of insect resistance engineered into tobacco, *Nature*, 300, 160, 1987.
595. **Hiatt, A., Cafferkey, R., and Bowdish, K.,** Production of antibodies in transgenic plants, *Nature*, 342, 76, 1989.
596. **Johnson, R., Narvaez, J., An, G., and Ryan, C.,** Expression of proteinase inhibitors I and II in transgenic tobacco plants: effects on natural defense against *Manduca sexta* larvae, *Proc. Natl. Acad. Sci U.S.A.*, 86, 9871, 1989.
597. **Yang, M. S., Espinoza, N. O., Nagpola, P. G., Dodds, J. H., White, F. F., Schnorr, K. L., and Jaynes, J. M.,** Expression of a synthetic gene for improved protein quality in transformed potato plants, *Plant Sci.*, 64, 99, 1989.
598. **Anthony-Cahill, S. J., Griffith, M. C., Noren, C. J., Suich, D. J., and Schultz, P. G.,** Site-specific mutagenesis with unnatural amino acids, *Trends Biochem. Sci.*, 14, 400, 1989.
599. **Bowie, J. U., Reidhaar-Olson, J. F., Wendell, A. L., and Sauer, R. T.,** Deciphering the message in protein sequences: tolerance to amino acid substitutions, *Science*, 247, 1306, 1990.
600. **Zoller, M. J.,** New recombinant DNA methodology for protein engineering, *Curr. Opin. Biotechnol.*, 3, 348, 1992.
601. **Medynski, D.,** Genetic approaches to protein structure and function: point mutations as modifiers of protein function, *Biotechnology*, 10, 1002, 1992.
602. **Wallace, J. C., Galili, G., Kawata, E. E., Cuellar, R. E., Shotwell, M. A., and Larkins, B. A.,** Aggregation of lysine-containing zeins into protein bodies in *Xenopus* oocytes, *Science*, 240, 662, 1988.
603. **Bagga, S., Sutton, D., Kemp, J. D., and Sengupta-Gopalan, C.,** Constitutive expression of the β-phaseolin gene in different tissues of transgenic alfalfa does not ensure phaseolin accumulation in non-seed tissue, *Plant Mol. Biol.*, 19, 951, 1992.
604. **Lawton, M. A., Tierney, M. A., Nakamura, I., Anderson, E., Komeda, Y., Dube, P., Hoffman, N., Fraley, R.T., and Beachy, R. N.,** Expression of a soybean β-conglycinin gene under the control of the cauliflower mosaic virus 35S and 19S promoters in transformed petunia tissues, *Plant Mol. Biol.*, 9, 315, 1987.
605. **Higgins, T. J. V. and Spencer, D.,** The expression of a chimeric cauliflower mosaic virus (CaMV-35S)-pea vicilin gene in tobacco, *Plant Sci.*, 74, 89, 1991.
606. **Ealing, P. M., Hancock, K. R., and White, D. W. R.,** Expression of the pea albumin 1 gene in transgenic white clover and tobacco, *Transgenic Res.*, 3, 344, 1994.
607. **Coraggio, I., Martegani, E., Compagno, C., Porro, D., Alberghina, L., Bernard, L., Faoro, F., and Viotti, A.,** Differential targeting and accumulation of normal and modified zein polypeptides in transformed yeast, *Eur. J. Cell Biol.*, 47, 165, 1988.
608. **Hellebust, H., Murby, M., Abrahmsen, L., Uhlen, M., and Enfors, S.-O.,** Different approaches to stabilize a recombinant fusion protein, *Biotechnology*, 7, 165, 1989.
609. **Rechsteiner, M., Rogers, S., and Rote, K.,** Protein structure and intracellular stability, *Trends Biochem. Sci.*, 12, 390, 1987.
610. **Mutter, M.,** Nature's rules and chemist's tools: a way for creating novel proteins, *Trends Biochem. Sci.*, 13, 260, 1988.
611. **Matsumura, M., Signor, G., and Matthews, B. W.,** Substantial increase of protein stability by multiple disulphide bonds, *Nature*, 342, 291, 1989.

612. Mahadevan, S., Erfle, J. D., and Sauer, F. D., J. Anim. Sci., 50, 723, 1980.
613. Hershko, A., Ubiquitin-mediated protein degradation, J. Biol. Chem., 263, 15237, 1988.
614. Monia, B. P., Ecker, D. J., and Crooke, S. T., New perspectives on the structure and function of ubiquitin, Biotechnology, 8, 209, 1990.
615. Butt, T. R., Jonnalagadda, S., Monia, B. P., Sternberg, E. J., Marsh, J., Stadel, J. M., Ecker, D. J., and Crooke, S. T., Ubiquitin fusion augments the yield of cloned gene products in Escherichia coli, Proc. Natl. Acad. Sci. U.S.A., 86, 2540, 1989.
616. Sabin, E. A., Lee-Ng, C. T., Shuster, J. R., and Barr, P. J., High-level expression and in vivo processing of chimeric ubiquitin fusion proteins in Saccharomyces cerevisae, Biotechnology, 7, 705, 1989.
617. Ecker, D. J., Stadel, J. M., Butt, T. R., Marsh, J. A., Monia, B. P., Powers, D. A., Gorman, J. A., Clark, P. E., Warren, F., Shatzman, A., and Crooke, S. T., Increasing gene expression in yeast by fusion to ubiquitin, J. Biol. Chem., 264, 7715, 1989.
618. Raikhel, N., Nuclear targeting in plants, Plant Physiol., 100, 1627, 1992.
619. McNulty, A. K. and Saunders, M. J., Purification and immunological detection of pea nuclear intermediate filaments: evidence for plant nuclear lamins, J. Cell Sci., 103, 407, 1992.
620. Tong, C.-G., Dauwalder, M., Clawson, G. A., Hatem, C. L., and Roux, S. J., The major nucleoside triphosphatase in pea (Pisum sativum L.) nuclei and in rat liver nuclei share common epitopes also present in nuclear lamins, Plant Physiol., 101, 1005, 1993.
621. Laskey, R. A. and Dingwall, C., Nuclear shuttling: the default pathway for nuclear proteins?, Cell, 74, 585, 1993.
622. Forbes, D. J., Structure and function of the nuclear pore complex, Annu. Rev. Cell Biol., 8, 495, 1992.
623. Dingwall, C. and Laskey, R., The nuclear membrane, Science, 258, 942, 1992.
624. Yamasaki, L. and Lanford, R. E., Nuclear transport: a guide to import receptors, Trends Cell Biol., 2, 123, 1992.
625. Goldfarb, D. S., Are the cytosolic components of the nuclear, ER, and mitochondrial import apparatus functionally related?, Cell, 70, 185, 1992.
626. Silver, P. A., How proteins enter the nucleus, Cell, 64, 489, 1991.
627. Nigg, E. A., Baeuerle, P. A., and Lührmann, R., Nuclear import-export: in search of signals and mechanisms, Cell, 66, 15, 1991.
628. Garcia-Bustos, J., Heitman, J., and Hall, M. N., Nuclear protein localization, Biochim. Biophys. Acta, 1071, 83, 1991.
629. Roberts, B., Nuclear location signal-mediated protein transport, Biochim. Biophys. Acta, 1008, 263, 1989.
630. Feldherr, C. M., Kallenbach, E., and Schultz, N., Movement of a karyophilic protein through the nuclear pores of oocytes, J. Cell Biol., 99, 2216, 1984.
631. Kalderon, D., Roberts, B. L., Richardson, W. D., and Smith, A. E., A short amino acid sequence able to specify nuclear location, Cell, 39, 499, 1984.
632. Kalderon, D., Richardson, W. D., Markham, A. F., and Smith, A. E., Sequence requirements for nuclear location of simian virus 40 large T antigen, Nature, 311, 33, 1984.
633. Robbins, J., Dilworth, S. M., Laskey, R. A., and Dingwall, C., Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence, Cell, 64, 615, 1991.
634. Fischer, U. and Lührmann, R., An essential signaling role for the m^{3}G cap in the transport of U1 snRNP to the nucleus, Science, 249, 786, 1990.
635. Michaud, N. and Goldfarb, D. S., Microinjected U snRNAs are imported to oocyte nuclei via the nuclear pore complex by three distinguishable targeting pathways, J. Cell Biol., 116, 851, 1992.
636. Adam, S. A., Marr, R. S., and Gerace, L., Nuclear protein import in permeabilized mammalian cells requires soluble cytoplasmic factors, J. Cell Biol., 111, 807, 1990.
637. Moore, M. S. and Blobel, G., The two steps of nuclear import, targeting to the nuclear envelope and translocation through the nuclear pore, require different cytosolic factors, Cell, 69, 939, 1992.
638. Stochaj, U., Osborne, M., Kurihara, T., and Silver, P., A yeast protein that binds nuclear localization sequences: purification, localization, and antibody inhibition of binding activity, J. Cell Biol., 113, 1243, 1991.
639. Adam, S. A. and Gerace, L., Cytosolic proteins that specifically bind nuclear localization signals are receptors for nuclear import, Cell, 66, 837, 1991.
640. Adam, E. J. H. and Adam, S. A., Identification of cytosolic factors required for nuclear location sequence-mediated binding to the nuclear envelope, J. Cell Biol., 125, 547, 1994.
641. Stochaj, U. and Silver, P., A conserved phosphoprotein that specifically binds nuclear localization sequences is essential for nuclear import, J. Cell Biol., 117, 473, 1992.
642. Wozniak, R. W., Blobel, G., and Rout, M. P., POM 152 is an integral protein of the pore membrane domain of the yeast nuclear envelope, J. Cell Biol., 125, 31, 1994.
643. Radu, A., Blobel, G., and Wozniak, R. W., Nup 155 is a novel nuclear pore complex protein that contains neither repetitive sequence motifs nor reacts with WGA, J. Cell Biol., 121, 1, 1993.
644. Shi, Y. and Thomas, J. O., The transport of proteins into the nucleus requires the 70-kilodalton heat shock protein or its cytosolic cognate, Mol. Cell. Biol., 12, 2186, 1992.
645. Standiford, D. M. and Richter, J. D., Analysis of a developmentally regulated nuclear localization signal in Xenopus, J. Cell Biol., 118, 991, 1992.

646. Pinol-Roma, S. and Dreyfuss, G., Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm, Nature, 355, 730, 1992.
647. Mandell, R. B. and Feldherr, C. M., Identification of two HSP 70-related Xenopus oocyte proteins that are capable of recycling across the nuclear envelope, J. Cell Biol., 111, 1775, 1990.
648. Meier, U. T. and Blobel, G., Nopp 140 shuttles on tracks between nucleolus and cytoplasm, Cell, 70, 127, 1992.
649. Schmidt-Zachmann, M. S., Dargemont, C., Kühn, L. C., and Nigg, E. A., Nuclear export of proteins: the role of nuclear retention, Cell, 74, 493, 1993.
650. Carrington, J. C., Freed, D. D., and Leinicke, A. J., Bipartite signal sequence mediates nuclear translocation of the plant potyviral NIa protein, Plant Cell, 3, 953, 1991.
651. Citovsky, V., Zupan, J., Warnick, D., and Zambryski, P., Nuclear localization of Agrobacterium VirE2 protein in plant cells, Science, 256, 1802, 1992.
652. Herrera-Estrella, A., Van Montagu, M., and Wang, K., A bacterial peptide acting as a plant nuclear targeting signal: the amino-terminal portion of Agrobacterium VirD2 protein directs a β-galactosidase fusion protein into tobacco nuclei, Proc. Natl. Acad. Sci. U.S.A., 87, 9534, 1990.
653. Howard, E. A., Zupan, J. R., Citovsky, V., and Zambryski, P. C., The VirD2 protein of A. tumefaciens contains a C-terminal bipartite nuclear localization signal: implications for nuclear uptake of DNA in plant cells, Cell, 68, 109, 1992.
654. Lassner, M. W., Jones, A., Daubert, S., and Comai, L., Targeting of T7 RNA polymerase to tobacco nuclei mediated by an SV40 nuclear location signal, Plant Mol. Biol., 17, 229, 1991.
655. Restrepo, M. A., Freed, D. D., and Carrington, J. C., Nuclear transport of plant potyviral proteins, Plant Cell, 2, 987, 1990.
656. Restrepo-Hartwig, M. A. and Carrington, J. C., Regulation of nuclear transport of a plant potyvirus protein by autoproteolysis, J. Virol., 66, 5662, 1992.
657. Van der Krol, A. R. and Chua, N.-H., The basic domain of plant B-ZIP proteins facilitates import of a reporter protein into plant nuclei, Plant Cell, 3, 667, 1991.
658. Varagona, M. J., Schmidt, R. J., and Raikhel, N. V., Monocot regulatory protein Opaque-2 is localized in the nucleus of maize endosperm and transformed tobacco plants, Plant Cell, 3, 105, 1991.
659. Varagona, M. J., Schmidt, R. J., and Raikhel, N. V., Nuclear localization signal(s) required for nuclear targeting of the maize regulatory protein, Opaque-2, Plant Cell, 4, 1213, 1992.
660. Varagona, M. J. and Raikhel, N. V., The basic domain in the bZIP regulatory protein Opaque-2 serves two independent functions: DNA binding and nuclear localization, Plant J., 5, 207, 1994.
661. Hicks, G. R. and Raikhel, N. V., Specific binding of nuclear localization sequences to plant nuclei, Plant Cell, 5, 983, 1993.
662. Shurvinton, C. E., Hodges, L., and Ream, W., A nuclear localization signal and the C-terminal omega sequence in the Agrobacterium tumefaciens VirD2 endonuclease are important for tumor formation, Proc. Natl. Acad. Sci. U.S.A., 89, 11837, 1992.
663. Tinland, B., Koukolíková-Nicola, Z., Hall, M. N., and Hohn, B., The T-DNA linked VirD2 protein contains two distinct functional nuclear localization signals, Proc. Natl. Acad. Sci. U.S.A., 89, 7442, 1992.
664. Koukolíková-Nicola, Z., Raineri, D., Stephens, K., Ramos, C., Tinland, B., Nester, E. W., and Hohn, B., Genetic analysis of the virD operon of Agrobacterium tumefaciens: a search for functions involved in the transport of T-DNA into the plant cell nucleus and in T-DNA integration, J. Bacteriol., 175, 723, 1993.
665. Li, X. H. and Carrington, J. C., Nuclear transport of tobacco etch potyviral RNA-dependent RNA polymerase is highly sensitive to sequence alterations, Virology, 193, 951, 1993.
666. Mikaélian, I., Drouet, E., Marechal, V., Denoyel, G., Nicolas, J.-C., and Sergeant, A., The DNA-binding domain of two bZIP transcription factors, the Epstein-Barr virus switch gene product EB1 and Jun, is a bipartite nuclear targeting sequence, J. Virol., 67, 734, 1993.
667. Shieh, M. W., Wessler, S. R., and Raikhel, N. V., Nuclear targeting of the maize R protein requiring two nuclear localization sequences, Plant Physiol., 101, 353, 1993.
668. Schmidt, R. J., Ketudat, M., Aukerman, M. J., and Hoschek, G., Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kDa zein genes, Plant Cell, 4, 689, 1992.
669. Trainer, I. and Verma, I. M., Identification of a nuclear targeting sequence in the Fos protein, Oncogene, 6, 2049, 1991.
670. Chida, K. and Vogt, P. K., Nuclear translocation of viral Jun but not of cellular Jun is cell cycle dependent, Proc. Natl. Acad. Sci. U.S.A., 89, 4290, 1992.
671. Ludwig, S. R., Habera, L. F., Dellaporta, S. L., and Wessler, S. R., Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region, Proc. Natl. Acad. Sci. U.S.A., 86, 7092, 1989.
672. Zambryski, P., Chronicles from the Agrobacterium-plant cell DNA transfer story, Annu. Rev. Plant Physiol. Plant Mol. Biol., 43, 465, 1992.
673. Olsen, L. J. and Harada, J. J., Biogenesis of peroxisomes in higher plants, in Molecular Approaches to Compartmentation and Metabolic Regulation, Huang, A. H. C. and Taiz, L., Eds., American Society of Plant Physiologists, Rockville, MD, 1991, 129.
674. Olsen, L. J., Ettinger, W. F., Damsz, B., Matsudaira, K., Webb, M. A., and Harada, J. J., Targeting of

glyoxysomal proteins to peroxisomes in leaves and roots of a higher plant, Plant *Cell*, 5, 941, 1993.
675. **Onyeocha, I., Behari, R., Hill, D., and Baker, A.,** Targeting of castor bean glyoxysomal isocitrate lyase to tobacco leaf peroxisomes, *Plant Mol. Biol.*, 22, 385, 1993.
676. **Beevers, H.,** Microbodies in higher plants, *Annu. Rev. Plant Physiol.*, 30, 159, 1979.
677. **Huang, A. H. C., Trelease, R. N., and Moore, T. S.,** *Plant Peroxisomes*, Academic Press, New York, 1983.
678. **Trelease, R. N.,** Biogenesis of glyoxysomes, *Annu. Rev. Plant Physiol.*, 35, 321, 1984.
679. **Tolbert, N. E.,** Metabolic pathways in peroxisomes and glyoxysomes, *Annu. Rev. Biochem.*, 50, 133, 1981.
680. **Newcomb, E. H.,** Ultrastructure and cytochemistry of plant peroxisomes and glyoxysomes, *Ann. NY Acad. Sci.*, 386, 228, 1982.
681. **Titus, D. E. and Becker, W. M.,** Investigation of the glyoxysome-peroxisome transition in germinating cucumber cotyledons using double-label immuno-electron microscopy, *J. Cell Biol.*, 101, 1288, 1985.
682. **Nishimura, M., Yamaguchi, J., Mori, H., Akazawa, T., and Yokota, S.,** Immunocytochemical analysis shows that glyoxysomes are directly transformed to leaf peroxisomes during greening of pumpkin cotyledons, *Plant Physiol.*, 80, 313, 1986.
683. **Sautter, C.,** Microbody transition in greening watermelon cotyledons. Double immunocytochemical labeling of isocitrate lyase and hydroxypyruvate reductase, *Planta*, 167, 491, 1986.
684. **Behrends, W., Birkhan, R., and Kindl, H.,** Transition form of microbodies. Overlapping of two sets of marker proteins during the rearrangement of glyoxysomes into leaf peroxisomes, *Biol. Chem. Hoppe-Seyler.*, 371, 85, 1990.
685. **De Bellis, L., Picciarelli, P., Pistelli, L., and Alpi, A.,** Localization of glyoxylate-cycle marker enzymes in peroxisomes of senescent leaves and green cotyledons, *Planta*, 180, 435, 1990.
686. **Subramani, S.,** Targeting of proteins into the peroxisomal matrix, *J. Memb. Biol.*, 125, 99, 1992.
687. **Gould, S. J., Keller, G.-A., and Subramani, S.,** Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase, *J. Cell Biol.*, 105, 2923, 1987.
688. **Gould, S. J., Keller, G.-A., Schneider, M., Howell, S. H., Garrard, L. J., Goodman, J. M., Distel, B., Tabak, H., and Subramani, S.,** Peroxisomal protein import is conserved between yeast, plants, insects and mammals, *EMBO J.*, 9, 85, 1990.
689. **Faber, K. N., Haima, P., DeHoop, M. J., Harder, W., Veenhuis, M., and Ab, G.,** Peroxisomal amine oxidase of *Hansenula polymorpha* does not require its SRL-containing C-terminal sequence for targeting, *Yeast*, 9, 331, 1993.
690. **Swinkels, B. W., Gould, S. J., Bodnar, A. G., Rachubinski, R. A., and Subramani, S.,** A novel, cleavable peroxisomal targeting signal at the amino-terminus of the rat 3-keto-acyl-CoA thiolase, *EMBO J.*, 10, 3255, 1991.
691. **Behari, R. and Baker, A.,** The carboxyterminus of isocitrate lyase is not essential for import into glyoxysomes in an *in vitro* system, *J. Biol. Chem.*, 268, 7315, 1993.
692. **Volokita, M.,** The carboxy-terminal end of glycolate oxidase directs a foreign protein into tobacco leaf peroxisomes, *Plant J.*, 1, 361, 1991.
693. **Gietl, C.,** Glyoxysomal malate dehydrogenase from watermelon is synthesized with an amino-terminal transit peptide, *Proc. Natl. Acad. Sci. U.S.A.*, 87, 5773, 1990.
694. **Van der Klei, I. J., Faber, K. N., Keizer-Gunnink, I., Gietl, C., Harder, W., and Veenhuis, M.,** Watermelon glyoxysomal malate dehydrogenase is sorted to peroxisomes of the methylotrophic yeast, *Hansenula polymorpha*, *FEBS Lett.*, 334, 128, 1993.
695. **Walton, P.A., Gould, S.J., Feramisco, J.R., and Subramani, S.,** Transport of microinjected proteins into peroxisomes of mammalian cells: inability of Zellweger cell lines to import proteins with the SKL tripeptide peroxisomal targeting signal, *Mol. Cell. Biol.*, 12, 531, 1992.
696. **Wolins, N. E. and Donaldson, R. P.,** Specific binding of the peroxisomal protein targeting sequence to glyoxysomal membranes, *J. Biol. Chem.*, 269, 1, 1994.
697. **Cuezva, J. M., Flores, A. I., Liras, A., Santarén, J. F., and Alconada, A.,** Molecular chaperones and the biogenesis of mitochondria and peroxisomes, *Biol. Cell*, 77, 47, 1993.
698. **Middalkoop, E., Strijland, A., and Tager, J. M.,** Does aminotriazole inhibit import of catalase into peroxisomes by retarding unfolding?, *FEBS Lett.*, 279, 79, 1991.
699. **Imanaka, T., Small, G. M., and Lazarow, P. B.,** Translocation of acyl-CoA oxidase into peroxisomes requires ATP hydrolysis but not a membrane potential, *J. Cell Biol.*, 105, 2915, 1987.
700. **Alvares, K., Carrillo, A., Yuan, P. M., Kawano, H., Morimoto, R. I., and Reddy, J. K.,** Identification of cytosolic peroxisome proliferator binding protein as a member of the heat shock protein HSP70 family, *Proc. Natl. Acad. Sci. U.S.A.*, 87, 5293, 1990.
701. **Smeekens, S., Weisbeek, P., and Robinson, C.,** Protein transport into and within chloroplasts, *Trends Biochem. Sci.*, 15, 73, 1990.
702. **Keegstra, K., Olsen, L. J., and Theg, S. M.,** Chloroplastic precursors and their transport across the envelope membranes, *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 40, 471, 1989.
703. **Weisbeek, P., Hageman, J., de Boer, D., and Smeekens, S.,** Transport of proteins towards the chloroplast thylakoid lumen, *Isr. J. Bot.*, 40, 123, 1991.
704. **Theg, S. M. and Scott, S. V.,** Protein import into chloroplasts, *Trends Cell Biol.*, 3, 186, 1993.
705. **Soll, J. and Alefsen, H.,** The protein import apparatus of chloroplasts, *Physiol. Plant.*, 87, 433, 1993.
706. **Robinson, C. and Klösgen, R. B.,** Targeting of proteins into and across the thylakoid membrane — a multitude of mechanisms, *Plant Mol. Biol.*, 26, 15, 1994.

707. **Ellis, J. R.,** Chloroplasts proteins: synthesis, transport and assembly, *Annu. Rev. Plant Physiol.*, 32, 111, 1981.
708. **Yuan, J., Cline, K., and Theg, S. M.,** Cryopreservation of chloroplasts and thylakoids for studies of protein import and integration, *Plant Physiol.*, 95, 1259, 1991.
709. **Robinson, C. and Ellis, R. J.,** Transport of proteins into chloroplasts. Partial purification of a chloroplast protease involved in the processing of imported precursor proteins, *Eur. J. Biochem.*, 142, 337, 1984.
710. **Robinson, C. and Ellis, R. J.,** Transport of proteins into chloroplasts. The precursor of small subunit of ribulose bisphosphate carboxylase in processed to the mature size in two steps, *Eur. J. Biochem.*, 142, 343, 1984.
711. **Mishkind, M. L., Wessler, S. R., and Schmidt, G. W.,** Functional determinants in transit sequences: import and partial maturation by vascular plant chloroplasts of the ribulose-1,5-bisphosphate carboxylase small subunit of *Chlamydomonas*, *J. Cell Biol.*, 100, 226, 1985.
712. **Smeekens, S., Van Steeg, H., Bauerle, C., Bettenbroek, H., Keegstra, K., and Weisbeek, P.,** Import into chloroplasts of a yeast mitochondrial protein directed by ferredoxin and plastocyanin transit peptides, *Plant Mol. Biol.*, 9, 377, 1987.
713. **Van den Broeck, G., Timko, M. P., Kausch, A. P., Cashmore, A. R., Van Montagu, M., and Herrera-Estrella, L.,** Targeting of a foreign protein to chloroplasts by fusion to the transit peptide from the small subunit of ribulose 1,5-bisphosphate carboxylase, *Nature*, 313, 358, 1985.
714. **Schreier, P. H., Seftor, E. A., Schell, J., and Bohnert, H. J.,** The use of nuclear-encoded sequences to direct the light-regulated synthesis and transport of a foreign protein into plant chloroplasts, *EMBO J.*, 4, 25, 1985.
715. **Lubben, T. H., Gatenby, A. A., Ahlquist, P., and Keegstra, K.,** Chloroplast import characteristics of chimeric proteins, *Plant Mol. Biol.*, 12, 13, 1989.
716. **Comai, L., Larson-Kelly, N., Kiser, J., Mau, C. J. D., Pokalsky, A. R., Shewmaker, C. K., McBride, K., Jones, A., and Stalker, D. M.,** Chloroplast transport of a ribulose bisphosphate carboxylase small subunit-5-enolpyruvyl 3-phosphoshikimate synthase chimeric protein requires part of the mature small subunit in addition to the transit peptide, *J. Biol. Chem.*, 263, 15104, 1988.
717. **Lubben, T. H. and Keegstra, K.,** Efficient *in vitro* import of a cytosolic heat shock protein into pea chloroplasts, *Proc. Natl. Acad. Sci. U.S.A.*, 83, 5502, 1986.
718. **Kavanagh, T. A., Jefferson, R. A., and Bevan, M. W.,** Targeting of a foreign protein to chloroplasts using fusions to the transit peptide of a chlorophyll a/b protein, *Mol. Gen. Genet.*, 215, 38, 1988.
719. **Meadows, J. W., Shackleton, J. B., Hulford, A., and Robinson, C.,** Targeting of a foreign protein into the thylakoid lumen of pea chloroplasts, *FEBS Lett.*, 253, 244, 1989.
720. **Ko, K. and Cashmore, A. R.,** Targeting of proteins to the thylakoid lumen by the bipartite transit peptide of the 33 kDa oxygen-evolving protein, *EMBO J.*, 8, 3187, 1989.
721. **Wasmann, C. C., Reiss, B., Bartlett, S. G., and Bohnert, H. J.,** The importance of the transit peptide and the transported protein for protein import into chloroplasts, *Mol. Gen. Genet.*, 205, 446, 1986.
722. **Della-Cioppa, G., Bauer, S. C., Taylor, M. L., Rochester, D. E., Klein, B. K., Shah, D. M., Fraley, R. T., and Kishore, G. M.,** Targeting a herbicide-resistant enzyme from *Escherichia coli* to chloroplasts of higher plants, *Biotechnology*, 5, 579, 1987.
723. **Von Heijne, G., Stepphun, J., and Herrmann, R. G.,** Domain structure of mitochondrial and chloroplast targeting peptides, *Eur. J. Biochem.*, 180, 535, 1989.
724. **Wasmann, C. C., Reiss, B., and Bohnert, H. J.,** Complete processing of a small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from pea requires the amino acid sequence Ile-Thr-Ser, *J. Biol. Chem.*, 263, 617, 1988.
725. **Reiss, B., Wasmann, C. C., and Bohnert, H. J.,** Regions in the transit peptide of SSU essential for transport into chloroplasts, *Mol. Gen. Genet.*, 209, 116, 1987.
726. **Reiss, B., Wasmann, C. C., Schell, J., and Bohnert, H. J.,** Effect of mutations on the binding and translocation functions of a chloroplast transit peptide, *Proc. Natl. Acad. Sci. U.S.A.*, 86, 886, 1989.
727. **Smeekens, S., Geerts, D., Bauerle, C., and Weisbeek, P.,** Essential function in chloroplast recognition of the ferredoxin transit peptide processing region, *Mol. Gen. Genet.*, 216, 178, 1989.
728. **Kuntz, M., Simons, A., Schell, J., and Schreier, P. H.,** Targeting of protein to chloroplasts in transgenic tobacco by fusion to mutated transit peptide, *Mol. Gen. Genet.*, 205, 454, 1986.
729. **Roise, D., Theiler, F., Horvath, S. J., Tomich, J. M., Richards, J. H., Allison, D. S., and Shatz, G.,** Amphiphilicity is essential for mitochondrial presequence function, *EMBO J.*, 7, 649, 1988.
730. **Von Heijne, G.,** Mitochondrial targeting sequences may form amphiphilic helices, *EMBO J.*, 5, 1335, 1986.
731. **Von Heijne, G. and Nishikawa, K.,** Hypothesis. Chloroplast transit peptides, the perfect random coil?, *FEBS Lett.*, 278, 1, 1991.
732. **Hageman, J., Baecke, C., Ebskamp, M., Pilon, R., Smeekens, S., and Weisbeek, P.,** Protein import into and sorting inside the chloroplast are independent processes, *Plant Cell*, 2, 479, 1990.
733. **Smeekens, S., Bauerle, C., Hageman, J., Keegstra, P., and Weisbeek, P.,** The role of the transit peptide in the routing of precursors toward different chloroplast compartments, *Cell*, 46, 365, 1986.
734. **Kohorn, B. D. and Tobin, E. M.,** A hydrophobic, carboxy-proximal region of a light-harvesting chlorophyll a/b protein is necessary for stable integration into thylakoid membranes, *Plant Cell*, 1, 159, 1989.

735. **Van den Broeck, G., Van Houtven, A., Van Montagu, M., and Herrera-Estrella, L.**, The transit peptide of a chlorophyll a/b-binding protein is not sufficient to insert neomycin phosphotransferase II in the thylakoid membrane, *Plant Sci.*, 58, 171, 1988.
736. **Huang, L., Adam, Z., and Hoffman, N. E.**, Deletion mutants of chlorophyll a/b binding proteins are efficiently imported into chloroplasts but do not integrate into thylakoid membranes, *Plant Physiol.*, 99, 247, 1992.
737. **Payan, L. A. and Cline, K.**, A stromal protein factor maintains the solubility and insertion competence of an imported thylakoid membrane protein, *J. Cell Biol.*, 112, 603, 1991.
738. **Yuan, J., Henry, R., and Cline, K.**, Stromal factor plays an essential role in protein integration into thylakoids that cannot be replaced by unfolding or by heat shock protein Hsp 70, *Proc. Natl. Acad. Sci. U.S.A.*, 90, 8552, 1993.
739. **Chitnis, P. R., Morishige, D. T., Nechushtai, R., and Thornber, J. P.**, Assembly of the barley light-harvesting chlorophyll a/b proteins in barley etiochloroplasts involves processing of the precursor on thylakoids, *Plant Mol. Biol.*, 11, 95, 1988.
740. **Broido, S., Loyter, A., and Vainstein, A.**, Transient expression of photosynthetic genes in transfected albinoid petunia protoplasts and correct processing of newly synthesized chloroplast-destined polypeptides, *Physiol. Plant.*, 88, 259, 1993.
741. **Yalovsky, S., Schuster, G., and Nechushtai, R.**, The apoprotein precursor of the major light-harvesting complex of photosystem II (LHCIIB) is inserted primarily into stromal lamellae and subsequently migrates to the grana, *Plant Mol. Biol.*, 14, 753, 1990.
742. **Cline, K., Ettinger, W. F., and Theg, S. M.**, Protein-specific energy requirements for protein transport across or into thylakoid membranes: two lumenal proteins are transported in the absence of ATP, *J. Biol. Chem.*, 267, 2688, 1992.
743. **Cai, D., Herrmann, R. G., and Klösgen, R. B.**, The 20 kDa apoprotein of the CP24 complex of photosystem II: an alternative model to study import and intraorganellar routing of nuclear-encoded thylakoid proteins, *Plant J.*, 3, 383, 1993.
744. **Smeekens, S. and Weisbeek, P.**, Protein transport towards the thylakoid lumen: post-translational translocation in tandem, *Photosyn. Res.*, 16, 177, 1988.
745. **Hageman, J., Robinson, C., Smeekens, S., and Weisbeek, P.**, A thylakoid processing protease is required for complete maturation of the lumen protein plastocyanin, *Nature*, 324, 567, 1986.
746. **Kirwin, P. M., Elderfield, P. D., and Robinson, C.**, Transport of proteins into chloroplasts. Partial purification of a thylakoidal processing peptidase involved in plastocyanin biogenesis, *J. Biol. Chem.*, 262, 16386, 1987.
747. **Kirwin, P. M., Elderfield, P. E., Williams, R. S., and Robinson, C.**, Transport of proteins into chloroplasts. Organization, orientation, and lateral distribution of the plastocyanin processing peptidase in the thylakoid network, *J. Biol. Chem.*, 263, 18128, 1988.
748. **James, H. E., Bartling, D., Musgrove, J. E., Kirwin, P. M., Herrmann, R. G., and Robinson, C.**, Transport of proteins into chloroplasts. Import and maturation of precursors to the 33-, 23-, and 16-kDa proteins of the oxygen-evolving complex, *J. Biol. Chem.*, 264, 19573, 1989.
749. **Howe, G. and Merchant, S.**, Maturation of thylakoid lumen proteins proceeds post-translationally through an intermediate *in vivo*, *Proc. Natl. Acad. Sci. U.S.A.*, 90, 1862, 1993.
750. **Konishi, T., Maruta, Y., Shinohara, K., and Watanabe, A.**, Transit peptides of thylakoid luminal proteins: the sites of stromal processing are conserved among higher plants, *Plant Cell Physiol.*, 34, 1081, 1993.
751. **Brock, I. W., Hazell, L., Michl, D., Nielsen, V. S., Moller, B. L., Herrmann, R. G., Klösgen, R. B., and Robinson, C.**, Precursors of one integral and five lumenal thylakoid proteins are imported by isolated pea and barley thylakoids: optimization of *in vitro* assays, *Plant Mol. Biol.*, 23, 717, 1993.
752. **Willey, D. L., Auffret, A. D., and Gray, J. C.**, Structure and topology of cytochrome f in pea chloroplast membranes, *Cell*, 36, 555, 1984.
753. **Salomon, M., Fischer, K., Flügge, U.-I., and Soll, J.**, Sequence analysis and protein import studies of an outer chloroplast envelope polypeptide, *Proc. Natl. Acad. Sci. U.S.A.*, 87, 5778, 1990.
754. **Li, H.-M., Moore, T., and Keegstra, K.**, Targeting of proteins to the outer envelope membrane uses a different pathway than transport into chloroplasts, *Plant Cell*, 3, 709, 1991.
755. **Flügge, U. I., Fischer, K., Gross, A., Sebald, W., Lottspeich, F., and Eckerskorn, C.**, The triose phosphate-3-phosphoglycerate-phosphate translocator from spinach chloroplasts: nucleotide sequence of a full-length cDNA clone and import of the *in vitro* synthesized precursor protein into chloroplasts, *EMBO J.*, 8, 39, 1989.
756. **Dreses-Werringloer, U., Fischer, K., Wachter, E., Link, T. A., and Flügge, U.-I.**, cDNA sequence and deduced amino acid sequence of the precursor of the 37-kDa inner envelope membrane polypeptide from spinach chloroplasts. Its transit peptide contains an amphiphilic α-helix as the only detectable structural element, *Eur. J. Biochem.*, 195, 361, 1991.
757. **Li, H.-M., Sullivan, T. D., and Keegstra, K.**, Information for targeting to the chloroplastic inner envelope membrane is contained in the mature region of the maize *Bt1*-encoded protein, *J. Biol. Chem.*, 267, 18999, 1992.
758. **Theg, S. M., Bauerle, C., Olsen, L. J., Selman, B. R., and Keegstra, K.**, Internal ATP is the only energy requirement for the translocation of precursor proteins across chloroplastic membranes, *J. Biol. Chem.*, 264, 6730, 1989.

759. **Olsen, L. J., Theg, S. M., Selman, B. R., and Keegstra, K.**, ATP is required for the binding of precursor proteins to chloroplasts, *J. Biol. Chem.*, 264, 6724, 1989.
760. **Olsen, L. J. and Keegstra, K.**, The binding of precursor proteins to chloroplasts requires nucleoside triphosphates in the intermembrane space, *J. Biol. Chem.*, 267, 433, 1992.
761. **Pain, D. and Blobel, G.**, Protein import into chloroplasts requires a chloroplast ATPase, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 3288, 1987.
762. **Cline, K.**, Import of proteins into chloroplasts: membrane integration of a thylakoid precursor protein reconstituted in chloroplast lysates, *J. Biol. Chem.*, 261, 14804, 1986.
763. **Viitanen, P. V., Doran, E. R., and Dunsmuir, P.**, What is the role of the transit peptide in thylakoid integration of the light-harvesting chlorophyll a/b protein?, *J. Biol. Chem.*, 263, 15000, 1988.
764. **Kirwin, P. M., Meadows, J. W., Shackleton, J. B., Musgrove, J. E., Elderfield, J. E., Mould, R., Hay, N. A., and Robinson, C.**, ATP-dependent import of a lumenal protein by isolated thylakoid vesicles, *EMBO J.*, 8, 2251, 1989.
765. **Konishi, T. and Watanabe, A.**, Transport of proteins into the thylakoid lumen: stromal processing and energy requirements for the import of the precursor to the 23 kDa protein of PSII, *Plant Cell Physiol.*, 34, 315, 1993.
766. **Robinson, C., Cai, D., Hulford, A., Brock, I. W., Michl, D., Hazell, L., Schmidt, I., Herrmann, R. G., and Klösgen, R. B.**, The presequence of a chimeric construct dictates which of two mechanisms is utilised for translocation across the thylakoid membrane: evidence for the existence of two distinct translocation systems, *EMBO J.*, 13, 279, 1994.
767. **Michl, D., Robinson, C., Shackleton, J. B., Herrmann, R. G., and Klösgen, R. B.**, Targeting of proteins to thylakoids by bipartite presequences: CFoII is imported by a novel, third pathway, *EMBO J.*, 13, 1310, 1994.
768. **Cline, K., Werner-Washburne, M., Lubben, T. H., and Keegstra, K.**, Precursors to two nuclear-encoded chloroplast proteins bind to the outer envelope membrane before being imported into chloroplasts, *J. Biol. Chem.*, 260, 3691, 1985.
769. **Friedman, A. L. and Keegstra, K.**, Chloroplast protein import: quantitative analysis of precursor binding, *Plant Physiol.*, 89, 993, 1989.
770. **Pfisterer, J., Lachmann, P., and Kloppstech, K.**, Transport of proteins into chloroplasts. Binding of nuclear-coded chloroplast proteins to the chloroplast envelope, *Eur. J. Biochem.*, 126, 143, 1982.
771. **Schnell, D. J. and Blobel, G.**, Identification of intermediates in the pathway of protein import into chloroplasts and their localization to envelope contact sites, *J. Cell Biol.*, 120, 103, 1993.
772. **Cornwell, K. L. and Keegstra, K.**, Evidence that a chloroplast surface protein is associated with a specific binding site for the precursor to the small subunit of ribulose-1,5-bisphosphate carboxylase, *Plant Physiol.*, 85, 780, 1987.
773. **Pain, D., Kanwar, Y. S., and Blobel, G.**, Identification of a receptor for protein import into chloroplasts and its localization to envelope contact zones, *Nature*, 331, 232, 1988.
774. **Joyard, J. and Douce, R.**, Import receptor in chloroplast envelope, *Nature*, 333, 306, 1988.
775. **Meyer, D. I.**, Mimics — or gimmicks?, *Nature*, 347, 424, 1990.
776. **Loddenkötter, B., Kammerer, B., Fischer, K., and Flügge, U.-I.**, Expression of the functional mature chloroplast triose phosphate translocator in yeast internal membranes and purification of the histidine-tagged protein by a single metal-affinity chromatography step, *Proc. Natl. Acad. Sci. U.S.A.*, 90, 2155, 1993.
777. **Hinz, G. and Flügge, U.-I.**, Phosphorylation of a 51-kDa envelope membrane polypeptide involved in protein translocation into chloroplasts, Eur. *J. Biochem.*, 175, 649, 1988.
778. **Waegemann, K. and Soll, J.**, Characterization of the protein import apparatus in isolated outer envelopes of chloroplasts, *Plant J.*, 1, 149, 1991.
779. **Soll, J. and Waegemann, K.**, A functionally active protein import complex from chloroplasts, *Plant J.*, 2, 253, 1992.
780. **Perry, S. E. and Keegstra, K.**, Envelope membrane proteins that interact with chloroplastic precursor proteins, *Plant Cell*, 6, 93, 1994.
781. **Amir-Shapira, D., Leustek, T., Dalie, B., Weissbach, H., and Brot, N.**, Hsp 70 proteins, similar to *Escherichia coli* DnaK, in chloroplasts and mitochondria of *Euglena gracilis*, *Proc. Natl. Acad. Sci. U.S.A.*, 87, 1749, 1990.
782. **Marshall, J. S., DeRocher, A. E., Keegstra, K., and Vierling, E.**, Identification of heat shock protein hsp70 homologues in chloroplasts, *Proc. Natl. Acad. Sci. U.S.A.*, 87, 374, 1990.
783. **Ko, K., Bornemisza, O., Kourtz, L., Ko, Z. W., Plaxton, W. C., and Cashmore, A.**, Isolation and characterization of a cDNA clone encoding a cognate 70-kDa heat shock protein of the chloroplast envelope, *J. Biol. Chem.*, 267, 2986, 1992.
784. **Marshall, J. S. and Keegstra, K.**, Isolation and characterization of a cDNA clone encoding the major Hsp70 of the pea chloroplast stroma, *Plant Physiol.*, 100, 1048, 1992.
785. **Wang, H., Goffreda, M., and Leustek, T.**, Characteristics of an Hsp 70 homolog localized in higher plant chloroplasts that is similar to DnaK, the Hsp 70 of prokaryotes, *Plant Physiol.*, 102, 843, 1993.
786. **Waegemann, K., Paulsen, H., and Soll, J.**, Translocation of proteins into isolated chloroplasts requires cytosolic factors to obtain import competence, *FEBS Lett.*, 261, 89, 1990.
787. **Guéra, A., America, T., van Waas, M., and Weisbeek, P. J.**, A strong protein unfolding activity

is associated with the binding of precursor chloroplast proteins to chloroplast envelopes, *Plant Mol. Biol.*, 23, 309, 1993.
788. **America, T., Hageman, J., Guéra, A., Rook, F., Archer, K., Keegstra, K., and Weisbeek, P.,** Methotrexate does not block import of a DHFR fusion protein into chloroplasts, *Plant Mol. Biol.*, 24, 283, 1994.
789. **Eilers, M. and Schatz, G.,** Binding of a specific ligand inhibits import of a purified precursor protein into mitochondria, *Nature*, 322, 228, 1986.
790. **Madueno, F., Napier, J. A., and Gray, J. C.,** Newly imported Rieske iron-sulfur protein associates with both Cpn 60 and Hsp 70 in the chloroplast stroma, *Plant Cell*, 5, 1865, 1993.
791. **Franklin, A. E. and Hoffman, N. E.,** Characterization of a chloroplast-homologue of the 54-kDa subunit of the signal recognition particle, *J. Biol. Chem.*, 268, 22175, 1993.
792. **Hemmingsen, S. M., Woolford, C., Van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W., and Ellis, J. R.,** Homologous plant and bacterial proteins chaperone oligomeric protein assembly, *Nature*, 333, 330, 1988.
793. **Roy, H.,** Rubisco assembly: a model system for studying the mechanism of chaperonin action, *Plant Cell*, 1, 1035, 1989.
794. **Ellis, R. J. and van der Vies, S.,** Molecular chaperones, *Annu. Rev. Biochem.*, 60, 321, 1991.
795. **Zabaleta, E., Assad, N., Oropeza, A., Salerno, G., and Herrera-Estrella, L.,** Expression of one of the members of the *Arabidopsis* chaperonin 60β gene family is developmentally regulated and wound-repressible, *Plant Mol. Biol.*, 24, 195, 1994.
796. **Fayet, O., Ziegelhoffer, T., and Georgopoulos, C. P.,** The GroES and GroEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures, *Bacteriology*, 171, 1379, 1989.
797. **Zweig, M. and Cummings, D. J.,** Cleavage of head and tail proteins during bacteriophage T5 assembly: selective host involvement in the cleavage of a tail protein, *J. Mol. Biol.*, 80, 505, 1973.
798. **Goloubinoff, P., Gatenby, A. A., and Lorimer, G. H.,** GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*, *Nature*, 337, 44, 1989.
799. **Bochkareva, E. S., Lissin, N. M., and Girshovich, A. S.,** Transient association of newly-synthesized unfolded proteins with the heat-shock GroEL protein, *Nature*, 336, 254, 1988.
800. **Lecker, S., Lill, R., Ziegelhoffer, T., Georgopoulos, C., Bassford, P. J., Kumamoto, C. A., and Wickner, W.,** Three pure chaperone proteins of *Escherichia coli*, *EMBO J.*, 8, 2703, 1989.
801. **Wu, H. B., Feist, G. L., and Hemmingsen, S. M.,** A modified *Escherichia coli* chaperonin (groEL) polypeptide synthesized in tobacco and targeted to the chloroplasts, *Plant Mol. Biol.*, 22, 1087, 1993.
802. **Grellet, F., Raynol, M., Laudié, M., Cooke, R., Giraudat, J., and Delseny, M.,** cDNA encoding a putative 10-kilodalton chaperonin from *Arabidopsis thaliana*, *Plant Physiol.*, 102, 685, 1993.
803. **Rintamaki, E.,** Rubisco subunit binding protein increases the solubility of rubisco large subunit *in vitro*, *Plant Physiol. Biochem.*, 29, 1, 1991.
804. **Lubben, T. H., Donaldson, G. K., Viitanen, P. V., and Gatenby, A. A.,** Several proteins imported into chloroplasts form stable complexes with the GroEL-related chloroplast molecular chaperone, *Plant Cell*, 1, 1223, 1989.
805. **Grimm, R., Donaldson, G. K., van der Vies, S. M., Schäfer, E., and Gatenby, A. A.,** Chaperonin-mediated reconstitution of the phytochrome photoreceptor, *J. Biol. Chem.*, 268, 5220, 1993.
806. **Dujon, B.,** Mitochondrial genetics and functions, in *Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*, Strathern, J., Jones, E., and Broach, J., Eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1982, 505.
807. **Attardi, G. and Schatz, G.,** Biogenesis of mitochondria, *Annu. Rev. Cell Biol.*, 4, 289, 1988.
808. **Hartl, F.-U., Pfanner, N., Nicholson, D. W., and Neupert, W.,** Mitochondrial protein import, *Biochim. Biophys. Acta*, 988, 1, 1989.
809. **Pfanner, N., Craig, E. A., and Meijer, M.,** The protein import machinery of the mitochondrial inner membrane, *Trends Biochem. Sci.*, 19, 368, 1994.
810. **Douglas, M. G., McCammon, M. T., and Vassarotti, A.,** Targeting proteins into mitochondria, *Microbiol. Rev.*, 50, 166, 1986.
811. **Hartl, F.-U. and Neupert, W.,** Protein sorting to mitochondria: evolutionary conservations of folding and assembly, *Science*, 247, 930, 1990.
812. **Ellis, R. J. and Robinson, C.,** Protein targeting, *Adv. Bot. Res.*, 14, 1, 1987.
813. **Glick, B. S., Beasley, E. M., and Shatz, G.,** Protein sorting in mitochondria, *Trends Biochem. Sci.*, 17, 453, 1992.
814. **Zimmermann, R., Paluch, U., Sprinzl, M., and Neupert, W.,** Cell-free synthesis of the mitochondrial ADP/ATP carrier protein of *Neurospora crassa*, *Eur. J. Biochem.*, 99, 247, 1979.
815. **Zimmermann, R., Paluch, U., and Neupert, W.,** Cell-free synthesis of cytochrome c, *FEBS Lett.*, 108, 141, 1979.
816. **Viebrock, A., Perz, A., and Sebald, W.,** The imported preprotein of the proteolipid subunit of the mitochondrial ATP synthase from *Neurospora crassa*. Molecular cloning and sequencing of the mRNA, *EMBO J.*, 1, 565, 1982.
817. **Kaput, J., Goltz, S., and Blobel, G.,** Nucleotide sequence of the yeast nuclear gene for cytochrome c peroxidase precursor. Functional implications of the pre-sequence for protein transport into mitochondria, *J. Biol. Chem.*, 257, 15054, 1982.
818. **Hurt, E. C., Pesold-Hurt, B., and Schatz, G.,** The cleavable prepiece of an imported mitochondrial protein is sufficient to direct cytosolic dihydrofolate reductase into the mitochondrial matrix, *FEBS Lett.*, 178, 306, 1984.
819. **Horwich, A. L., Kalousek, F., Mellman, I., and Rosenberg, L. E.,** A leader peptide is sufficient to direct mitochondrial import of a chimeric protein, *EMBO J.*, 4, 1129, 1985.
820. **Boutry, M., Nagy, F., Poulsen, C., Aoyagi, K., and Chua, N.-H.,** Targeting of bacterial chloramphenicol acetyltransferase to mitochondria in transgenic plants, *Nature*, 328, 340, 1987.
821. **Hemon, P., Robbins, M. P., and Cullimore, J. V.,** Targeting of glutamine synthetase to the mitochondria of transgenic tobacco, *Plant Mol. Biol.*, 15, 895, 1990.
822. **Kimura, T., Takeda, S., Kyozuka, J., Asahi, T., Shimamoto, K., and Nakamura, K.,** The presequence of a precursor to the δ-subunit of sweet potato mitochondrial F_1ATPase is not sufficient for the transport of β-glucuronidase (GUS) into mitochondria of tobacco, rice and yeast cells, *Plant Cell Physiol.*, 34, 345, 1993.
823. **White, J. A. and Scandalios, J. G.,** Deletion analysis of the maize mitochondrial superoxide dismutase transit peptide, *Proc. Natl. Acad. Sci. U.S.A.*, 86, 3534, 1989.
824. **Schmitz, U. K. and Lonsdale, D. M.,** A yeast mitochondrial presequence functions as a signal for targeting to plant mitochondria *in vivo*, *Plant Cell*, 1, 783, 1989.
825. **Huang, J., Hack, E., Thornburg, R. W., and Myers, A. M.,** A yeast mitochondrial leader peptide functions *in vivo* as a dual targeting signal for both chloroplasts and mitochondria, *Plant Cell*, 2, 1249, 1990.
826. **Roise, D. and Shatz, G.,** Mitochondrial presequences, *J. Biol. Chem.*, 263, 4509, 1988.
827. **Eisenberg, D., Weiss, R. M., and Terwilliger, T. C.,** The hydrophobic moment detects periodicity in protein hydrophobicity, *Proc. Natl. Acad. Sci. U.S.A.*, 81, 140, 1984.
828. **Kaiser, E. T. and Kezdy, F. J.,** Peptides with affinity for membranes, *Annu. Rev. Biophys. Biophys. Chem.*, 16, 561, 1987.
829. **Schleyer, M. and Neupert, W.,** Transport of proteins into mitochondria: translocation intermediates spanning contact sites between outer and inner membranes, *Cell*, 43, 339, 1985.
830. **Pfanner, N., Hartl, F.-U., Guiard, B., and Neupert, W.,** Mitochondrial precursor proteins are imported through a hydrophilic membrane environment, *Eur. J. Biochem.*, 169, 289, 1987.
831. **Pfanner, N., Rassow, J., Van der Klei, I. J., and Neupert, W.,** A dynamic model of the mitochondrial protein import machinery, *Cell*, 68, 999, 1992.
832. **Bohni, P., Gasser, S., Leaver, C., and Schatz, G.,** A matrix-localized mitochondrial protease processing cytoplasmically made precursors to mitochondrial proteins, in *The Organization and Expression of the Mitochondrial Genome*, Kroon, A. M. and Saccone, C., Eds., Elsevier, North Holland, Amsterdam, 1980, 323.
833. **Witte, C., Jensen, R. E., Yaffe, M. P., and Schatz, G.,** *MAS1*, a gene essential for yeast mitochondrial assembly, encodes a subunit of the mitochondrial processing protease, *EMBO J.*, 7, 1439, 1988.
834. **Hawlitschek, G., Schneider, H., Schmidt, B., Tropschug, M., Hartl, F.-U., and Neupert, W.,** Mitochondrial protein import: identification of processing peptidase and of PEP, a processing enhancing protein, *Cell*, 53, 795, 1988.
835. **Kalousek, F., Neupert, W., Omura, T., Schatz, G., and Schmitz, U. K.,** Uniform nomenclature for the mitochondrial peptidases cleaving precursors of mitochondrial proteins, *Trends Biochem. Sci.*, 18, 249, 1993.
836. **Sztul, E. S., Chu, T. W., Strauss, A. W., and Rosenberg, L. E.,** Import of the malate dehydrogenase precursor by mitochondria. Cleavage within leader peptide by matrix protease leads to formation of intermediate-sized form, *J. Biol. Chem.*, 263, 12085, 1988.
837. **Jensen, R. E. and Yaffe, M. P.,** Import of proteins into yeast mitochondria: the nuclear *MAS2* gene encodes a component of the processing protease that is homologous to the *MAS1*-encoded subunit, *EMBO J.*, 7, 3863, 1988.
838. **Yang, M., Jensen, R. E., Yaffe, M. P., Oppliger, W., and Schatz, G.,** Import of proteins into yeast mitochondria: the purified matrix processing protease contains two subunits which are encoded by the nuclear *MAS 1* and *MAS 2* genes, *EMBO J.*, 7, 3857, 1988.
839. **Arretz, M., Schneider, H., and Neupert, W.,** Processing of mitochondrial precursor proteins, *Biomed. Biochem. Acta*, 50, 403, 1991.
840. **Braun, H. P., Emmermann, M., Kruft, V., and Schmitz, U. K.,** The general processing peptidase from potato is an integral part of cytochrome c reductase of the respiratory chain, *EMBO J.*, 11, 3219, 1992.
841. **Eriksson, A. C., Sjöling, S., and Glaser, E.,** The ubiquinol cytochrome c oxidoreductase complex of spinach leaf mitochondria is involved in both respiration and protein processing, *Biochim. Biophys. Acta*, 1186, 221, 1994.
842. **Eriksson, A. C. and Glaser, E.,** Mitochondrial processing proteinase: a general processing proteinase of spinach leaf mitochondria is a membrane-bound enzyme, *Biochim. Biophys. Acta*, 1140, 208, 1992.
843. **Knorpp, C., Hugosson, M., Sjöling, S., Eriksson, A. C., and Glaser, E.,** Tissue-specific differences of the mitochondrial protein import machinery: *in vitro* import, processing and degradation of the pre-$F_1\beta$ subunit of the ATP synthase in spinach leaf and root mitochondria, *Plant Mol. Biol.*, 26, 571, 1994.
844. **Guiard, B.,** Structure, expression and regulation of a nuclear gene encoding a mitochondrial protein: the yeast L(+)-lactate cytochrome c oxido-reductase (cytochrome b_2), *EMBO J.*, 4, 3265, 1985.

845. Römisch, J., Tropschug, M., Sebald, W., and Weiss, H., The primary structure of cytochrome c1 from *Neurospora crassa*, *Eur. J. Biochem.*, 164, 111, 1987.
846. Sadler, I., Suda, K., Schatz, G., Kaudewitz, F., and Haid, A., Sequencing of the nuclear gene for the yeast cytochrome c_1 precursor reveals an unusually complex amino-terminal presequence, *EMBO J.*, 3, 2137, 1984.
847. Hartl, F.-U., Ostermann, J., Guiard, B., and Neupert, W., Successive translocation into and out of the mitochondrial matrix: targeting of proteins to the intermembrane space by a bipartite signal peptide, *Cell*, 51, 1027, 1987.
848. Van Loon, A. P. G. M., Brandli, A. W., Pesold-Hurt, B., Blank, D., and Schatz, G., Transport of proteins to the mitochondrial intermembrane space: the "matrix-targeting" and the "sorting" domains in the cytochrome c_1 presequence, *EMBO J.*, 6, 2433, 1987.
849. Van Loon, A. P. G. M. and Schatz, G., Transport of proteins to the mitochondrial intermembrane space: the "sorting" domain of the cytochrome c_1 presequence is a stop-transfer sequence specific for the mitochondrial inner membrane, *EMBO J.*, 6, 2441, 1987.
850. Gasser, S., Ohashi, A., Daum, G., Bohni, P., Gibson, J., Reid, G., Yonetani, T., and Schatz, G., Imported mitochondrial proteins cytochrome b_2 and cytochrome c_1 are processed in two steps, *Proc. Natl. Acad. Sci. U.S.A.*, 79, 267, 1982.
851. Teintze, M., Slaughter, M., Weiss, H., and Neupert, W., Biogenesis of mitochondrial ubiquinol: cytochrome c reductase (cytochrome bc_1 complex), *J. Biol. Chem.*, 257, 10364, 1982.
852. Hatefi, Y., The mitochondrial electron transport and oxidative phosphorylation system, *Annu. Rev. Biochem.*, 54, 1015, 1985.
853. Hartl, F.-U., Schmidt, B., Wachter, E., Weiss, H., and Neupert, W., Transport into mitochondria and intramitochondrial sorting of Fe/S protein of ubiquinol-cytochrome c reductase, *Cell*, 47, 939, 1986.
854. Harnisch, U., Weiss, H., and Sebald, W., The primary structure of the iron-sulfur subunit of ubiquinol-cytochrome c reductase from *Neurospora*, determined by cDNA and gene sequencing, *Eur. J. Biochem.*, 149, 95, 1985.
855. Beckman, J. D., Ljungdahl, P. O., Lopez, J. L., and Trumpower, B. L., Isolation and characterization of the nuclear gene encoding the Reiske iron-sulfur protein (RIP1) from *Saccharomyces cerevisiae*, *J. Biol. Chem.*, 262, 8901, 1987.
856. Emmermann, M., Clericus, M., Braun, H.-P., Mozo, T., Heins, L., Kruft, V., and Schmitz, U. K., Molecular features, processing and import of the Rieske iron-sulfur protein from potato mitochondria, *Plant Mol. Biol.*, 25, 271, 1994.
857. Pfanner, N., Hoeben, P., Tropschug, M., and Neupert, W., The carboxy-terminal two-thirds of the ADP/ATP carrier polypeptide contains sufficient information to direct translocation into mitochondria, *J. Biol. Chem.*, 262, 14851, 1987.
858. Nicholson, D. W. and Neupert, W., Import of cytochrome c into mitochondria: reduction of heme, mediated by NADH and flavin nucleotides, is obligatory for its covalent linkage to apocytochrome c, *Proc. Natl. Acad. Sci. U.S.A.*, 86, 4340, 1989.
859. Stuart, R. A., Nicholson, D. W., and Neupert, W., Early steps in mitochondrial protein import: receptor functions can be substituted by the membrane insertion activity of apocytochrome c, *Cell*, 60, 31, 1990.
860. Schwaiger, M., Herzog, V., and Neupert, W., Characterization of translocation contact sites involved in the import of mitochondrial proteins, *J. Cell Biol.*, 105, 235, 1987.
861. Vestweber, D. and Schatz, G., Point mutations destabilizing a precursor protein enhance its post-translational import into mitochondria, *EMBO J.*, 7, 1147, 1988.
862. Rassow, J., Guiard, B., Wienhues, U., Herzog, V., Hartl, F.-U., and Neupert, W., Translocation arrest by reversible folding of a precursor protein imported into mitochondria: a means to quantitate translocation contact sites, *J. Cell Biol.*, 109, 1412, 1989.
863. Zimmermann, R., Sagstetter, M., Lewis, M. J., and Pelham, H. R. B., Seventy-kilodalton heat shock proteins and an additional component from reticulocyte lysate stimulate import of M13 procoat protein into microsomes, *EMBO J.*, 7, 2875, 1988.
864. Stuart, R. A., Cyr, D. M., Craig, E. A., and Neupert, W., Mitochondrial molecular chaperones: their role in protein translocation, *Trends Biochem. Sci.*, 19, 87, 1994.
865. Lithgow, T., Ryan, M., Anderson, R. L., Hoj, P. B., and Hoogenraad, N. J., A constitutive form of heat-shock protein 70 is located in the outer membranes of mitochondria from rat liver, *FEBS Lett.*, 332, 277, 1993.
866. Miernyk, J. A., Duck, N. B., David, N. R., and Randall, D. D., Autophosphorylation of the pea mitochondrial heat-shock protein homolog, *Plant Physiol.*, 100, 965, 1992.
867. Kang, P.-J., Ostermann, J., Shilling, J., Neupert, W., Craig, E. A., and Pfanner, N., Requirement for hsp70 in the mitochondrial matrix for translocation and folding of precursor proteins, *Nature*, 348, 137, 1990.
868. Murakami, H., Pain, D., and Blobel, G., 70-kDa heat shock-related protein is one of at least two distinct cytosolic factors stimulating protein import into mitochondria, *J. Cell Biol.*, 107, 2051, 1988.
869. Randall, S. K. and Shore, G. C., Import of a mutant mitochondrial precursor fails to respond to stimulation by a cytosolic factor, *FEBS Lett.*, 250, 561, 1989.
870. Kiebler, M., Keil, P., Schneider, H., van der Klei, I. J., Pfanner, N., and Neupert, W., The mitochondrial receptor complex: a central role of MOM 22 in mediating preprotein transfer from receptors to the general insertion pore, *Cell*, 74, 483, 1993.
871. Pfaller, R. and Neupert, W., High-affinity binding sites involved in the import of porin into mitochondria, *EMBO J.*, 6, 2635, 1987.

872. Pfaller, R., Pfanner, N., and Neupert, W., Mitochondrial protein import: bypass of proteinaceous surface receptors can occur with low specificity and efficiency, *J. Biol. Chem.*, 264, 34, 1989.
873. Sollner, T., Griffith, G., Pfaller, R., Pfanner, N., and Neupert, W., MOM 19, an import receptor for mitochondrial precursor proteins, *Cell*, 59, 1061, 1989.
874. Pfanner, N., Tropschug, M., and Neupert, W., Mitochondrial protein import: nucleoside triphosphates are involved in conferring import-competence to precursors, *Cell*, 49, 815, 1987.
875. Gambill, B.D., Voos, W., Kang, P. J., Miao, B., Langer, T., Craig, E. A., and Pfanner, N., A dual role for mitochondrial heat shock protein 70 in membrane translocation of preproteins, *J. Cell Biol.*, 123, 109, 1993.
876. Voos, W., Gambill, B. D., Guiard, B., Pfanner, N., and Craig, E. A., Presequence and mature part of preproteins strongly influence the dependence of mitochondrial protein import on heat shock protein 70 in the matrix, *J. Cell Biol.*, 123, 119, 1993.
877. Manning-Krieg, U. C., Scherer, P. E., and Schatz, G., Sequential action of mitochondrial chaperones in protein import into the matrix, *EMBO J.*, 10, 3273, 1991.
878. Martin, J., Mahlke, K., and Pfanner, N., Role of an energized inner membrane in mitochondrial protein import. $\Delta\Psi$ drives the movement of presequences, *J. Biol. Chem.*, 266, 18051, 1991.
879. Cheng, M. Y., Hartl, F.-U., Pollock, R. A., Kalousek, F., Neupert, W., Hallberg, E. M., Hallberg, R. L., and Horwich, A. L., Mitochondrial heat-shock protein HSP60 is essential for assembly of proteins imported into yeast mitochondria, *Nature*, 337, 620, 1989.
880. Ostermann, J., Horwich, A. L., Neupert, W., and Hartl, F.-U., Protein folding in mitochondria requires complex formation with hsp 60 and ATP hydrolysis, *Nature*, 341, 125, 1989.
881. Reading, D. S., Hallberg, R. L., and Myers, A. M., Characterization of the yeast *HSP60* gene coding for a mitochondrial assembly factor, *Nature*, 337, 655, 1989.
882. Franz, J. E., Discovery, development and chemistry of glyphosate, in *The Herbicide Glyphosate*, Grossman, E. and Atkinson, D., Eds., Butterworths, London, 1985, 3.
883. Steinrucken, H. C. and Amrhein, N., The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase, *Biochem. Biophys. Res. Commun.*, 94, 1207, 1980.
884. Shah, D. M., Horsch, R. B., Klee, H. J., Kishore, G. M., Winter, J. A., Tumer, N. E., Hironaka, C. M., Sanders, P. R., Gasser, C. S., Aykent, S., Siegal, N. R., Rogers, S. G., and Fraley, R. T., Engineering herbicide tolerance in transgenic plants, *Science*, 233, 478, 1986.
885. Nagley, P. and Devenish, R. J., Leading organellar proteins along new pathways: the relocation of mitochondrial and chloroplast genes to the nucleus, *Trends Biochem. Sci.*, 14, 31, 1989.
886. Vermaas, W. F. J., Arntzen, C. J., Gu, L.-Q., and Yu, C. A., Interactions of herbicides and azidoquinones at a photosystem II binding site in the thylakoid membrane, *Biochim. Biophys. Acta*, 723, 266, 1983.
887. Hirschberg, J. and McIntosh, L., Molecular basis of herbicide resistance in *Amaranthus hybridus*, *Science*, 222, 1346, 1984.
888. Cheung, A. Y., Bogorad, L., Van Montagu, M., and Schell, J., Relocating a gene for herbicide tolerance: a chloroplast gene is converted into a nuclear gene, *Proc. Natl. Acad. Sci. U.S.A.*, 85, 391, 1988.
889. Stitt, M., Control of photosynthetic carbon fixation and partitioning: how can use of genetically manipulated plants improve the nature and quality of information about regulation?, *Philos. Trans. R. Soc. Lond. B*, 342, 225, 1993.
890. Bradley, D., van der Vies, S. M., and Gatenby, A. A., Expression of cyanobacterial and higher plant ribulose 1,5-bisphosphate carboxylase genes in *Escherichia coli*, *Philos. Trans. R. Soc. Lond. B*, 313, 447, 1986.
891. Winter, P. and Herrmann, R. G., *Bot. Acta*, 101, 42, 1987.
892. Nagley, P., Farrell, L. B., Gearing, D. P., Nero, D., Meltzer, S., and Devenish, R. J., Assembly of functional proton-translocating ATPase complex in yeast mitochondria with cytoplasmically synthesized subunit 8, a polypeptide normally encoded within the organelle, *Proc. Natl. Acad. Sci. U.S.A.*, 85, 2091, 1988.
893. Banroques, J., Délahodde, A., and Jacq, C., A mitochondrial RNA maturase gene transferred to the yeast nucleus can control mitochondrial mRNA slicing, *Cell*, 46, 837, 1986.
894. Howe, C. J., Organelle transformation, *Trends Genet.*, 4, 150, 1988.
895. Vestweber, D. and Schatz, G., DNA-protein conjugates can enter mitochondria via the protein import pathway, *Nature*, 338, 170, 1989.
896. O'Neill, C., Horváth, G. V., Horváth, E., Dix, P. J., and Medgyesy, P., Chloroplast transformation in plants: polyethylene glycol (PEG) treatment of protoplasts is an alternative to biolistic delivery systems, *Plant J.*, 3, 729, 1993.
897. Maliga, P., Carrer, H., Kanevski, I., Staub, J., and Svab, Z., Plastid engineering in land plants: a conservative genome is open to change, *Philos. Trans. R. Soc. Lond. B.*, 342, 203, 1993.
898. Maliga, P., Toward plastid transformation in flowering plants, *TIBTECH*, 11, 101, 1993.
899. Marrison, J. L. and Leech, R. M., The subcellular and intra-organelle recognition of nuclear and chloroplast transcripts in developing leaf cells, *Plant J.*, 6, 605, 1994.